



WOMEN IN DATA





El material de esta presentación se encuentra disponible en Google Colab
Haz click en el logo de arriba



<https://alkzar.cl>



@vamos_alcazar





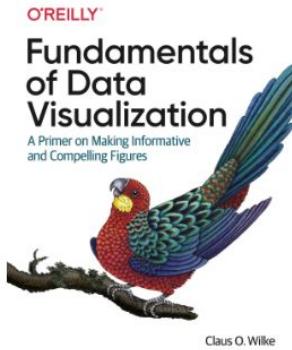
- + Se asocian las visualizaciones a sus funciones respectivas para crearlas
- + Modificación de algunos parámetros
- + Crear visualizaciones por capas
- + Uso de anotaciones para explicar y guiar
- + Implicancia de los principales parámetros y sus modificaciones

**“Chart design, like writing, is
as much a science as it is an art.”**

Alberto Cairo, **How Charts Lie: Getting Smarter about Visual Information**

Contenido de la presentación

‣ Capítulo 1: Visualizar cantidades



Implementación en python [2](#), de varias de las buenas prácticas desarrolladas en el capítulo 6: *Visualizing Amounts* del libro [Fundamentals of Data Visualization \(Wilke\)](#).

¿Qué encontrarás en esta sección?

1. Lo esencial de visualizar magnitudes usando gráficos de barra
2. Agregar como axes de un subplot visualizaciones creadas con `seaborn`
3. Utilizar un heatmap para ver la evolución de magnitudes a lo largo del tiempo
4. La importancia del orden en revelar las tendencias

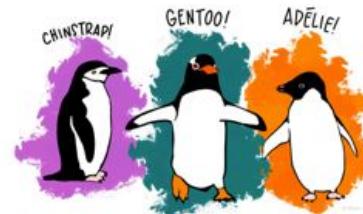
↵ 69 cells hidden



Contenido de la presentación

▶ Capítulo 1: Visualizar car

▶ Capítulo 2: Palmer Penguins



Utilizaremos el dataset [Palmer Penguins](#) disponible en el repositorio por @allison_horst.

¿Qué encontrarás en esta sección? 🔎

Implementación en python [2](#), de v:
[Fundamentals of Data Visualization](#)

¿Qué encontrarás en esta sección?,

1. Lo esencial de visualizar mag
2. Agregar como axes de un sub
3. Utilizar un heatmap para ver l
4. La importancia del orden en r

↳ 69 cells hidden

↳ 10 cells hidden

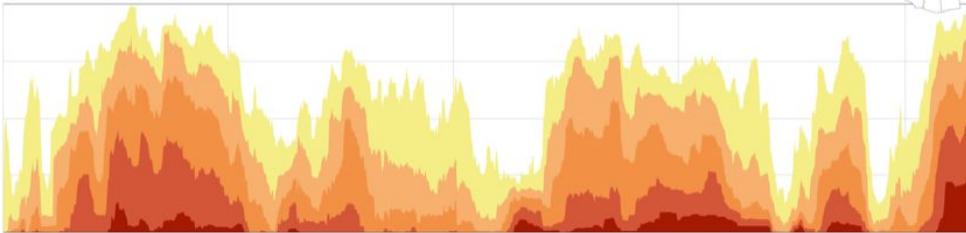


Contenido de la presentación

‣ Capítulo 3: U.S. Drought

‣ Capítulo 1: Visualizar car

‣ Capítulo 2: Palmer Pe



Utilizaremos el dataset [Palme](#)

¿Qué encontrarás en esta sección?

1. Crear una visualización con [matplotlib](#)
2. Modificar sus elementos
3. Detalles de edición en matplotlib:
 - Agregar anotaciones
 - Agregar una imagen

¿Qué encontrarás en esta sección? ↗

1. Preparar los datos para la visualización con [pandas](#)
2. Reproducir el *stacked area chart* publicado en el artículo del NYC usando [matplotlib](#) (la visualización de arriba)
3. La importancia en la elección de los colores y algunos *antipatterns*
4. Detalles de edición en [matplotlib](#) como:
 - Agregar etiquetas custom
 - Cambiar los breaks y ticks del eje y
 - Modificar las etiquetas del eje x

Implementación en python de [DataCamp](#): [Fundamentals of Data Visualization](#)

¿Qué encontrarás en esta sección?,

1. Lo esencial de visualizar mag
2. Agregar como axes de un sub
3. Utilizar un heatmap para ver lo
4. La importancia del orden en ru

↳ 69 cells hidden

↳ 10 cells hidden

↳ 36 cells hidden



#1

Visualizar cantidades

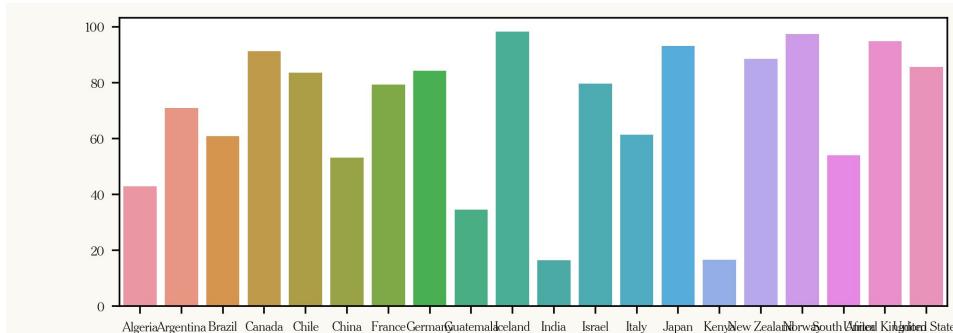


Invertir ejes
etiquetas más legibles

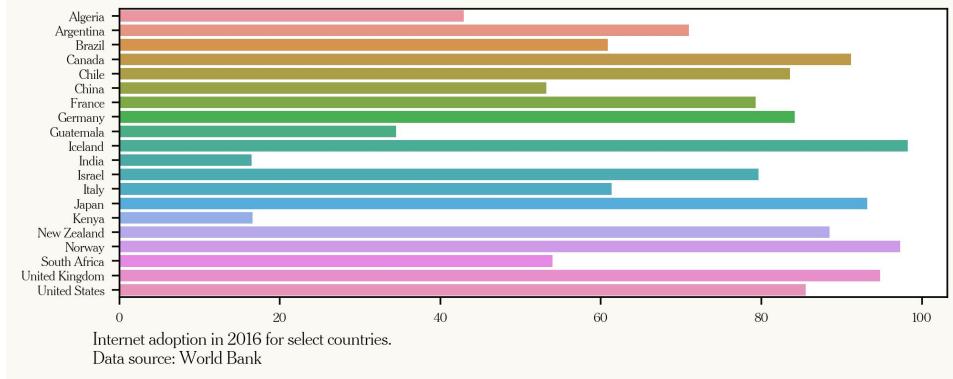
Niveles similares, contrastar las
diferencias

De múltiples barras-a-heatmaps

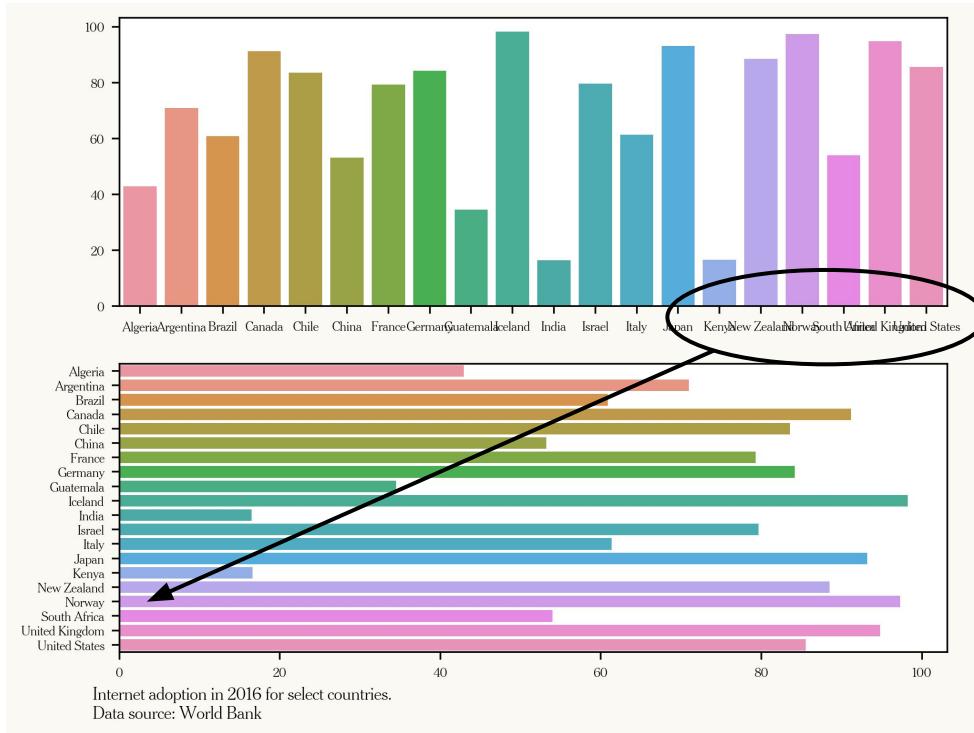
1.A Superposición de etiquetas: problema común con número moderado de categorías y etiquetas descriptivas



```
sns.barplot(x='country_name',
             y='value',
             data=internet2016,
             ax=ax[0],
             ci=False)
```



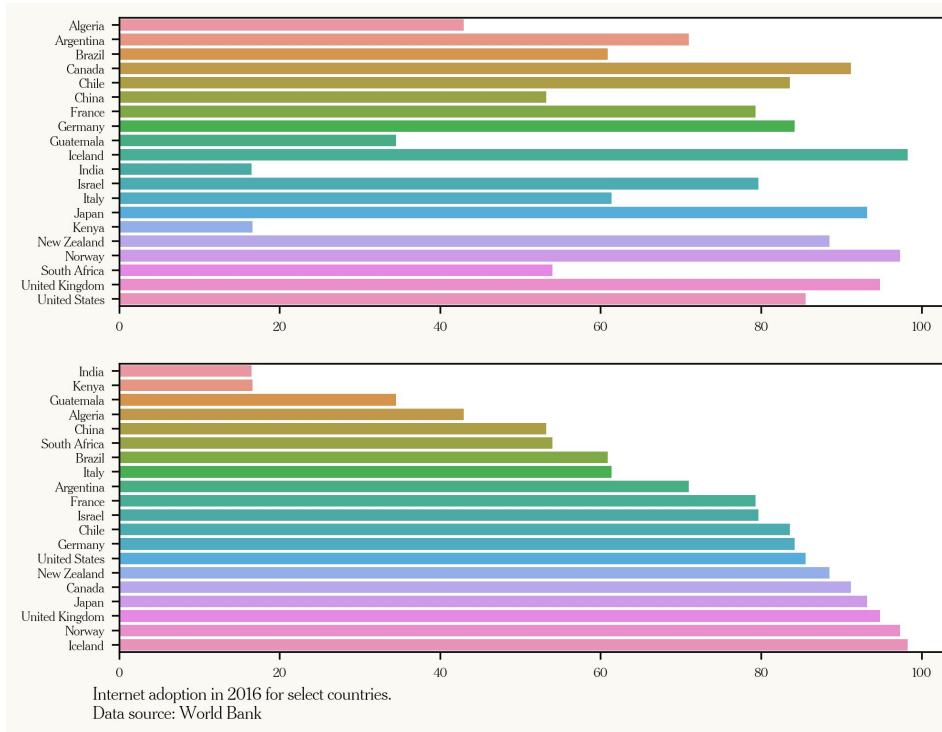
1.A Invertir coordenadas: solución a la superposición de etiquetas y mejora en la legibilidad de la visualización



```
sns.barplot(x='country_name',
             y='value',
             data=internet2016,
             ax=ax[0],
             ci=False)
```

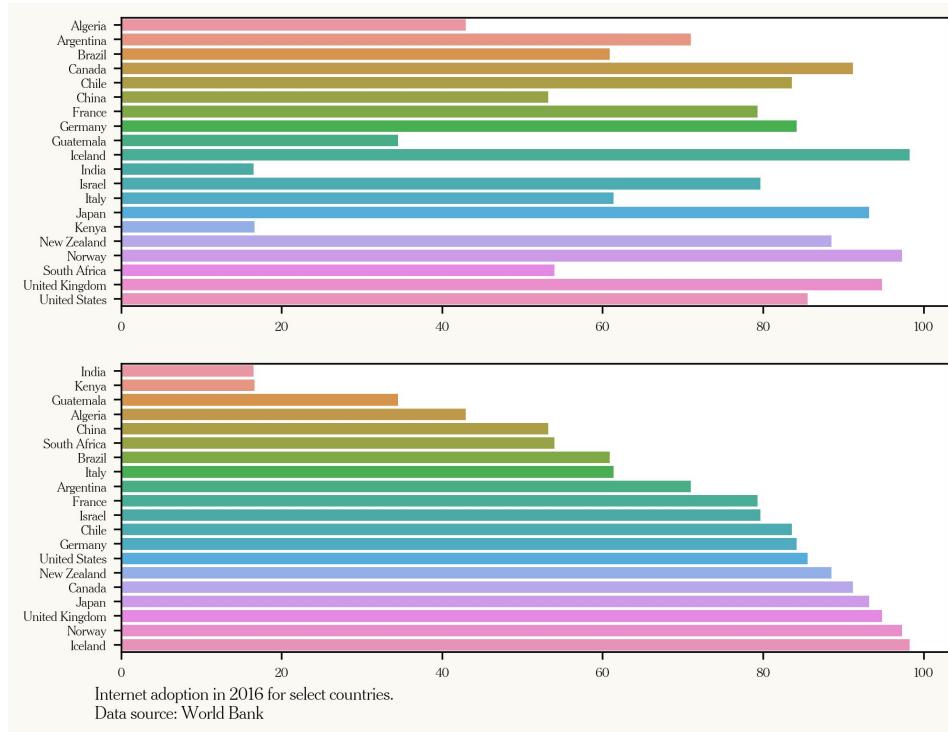
```
sns.barplot(x='value',
             y='country_name',
             data=internet2016,
             ax=ax[1],
             ci=False)
```

1.B Orden alfabético: no tiene relación con las magnitudes de las categorías



```
sns.barplot(x='country_name',
             y='value',
             data=internet2016,
             ax=ax[0],
             ci=False)
```

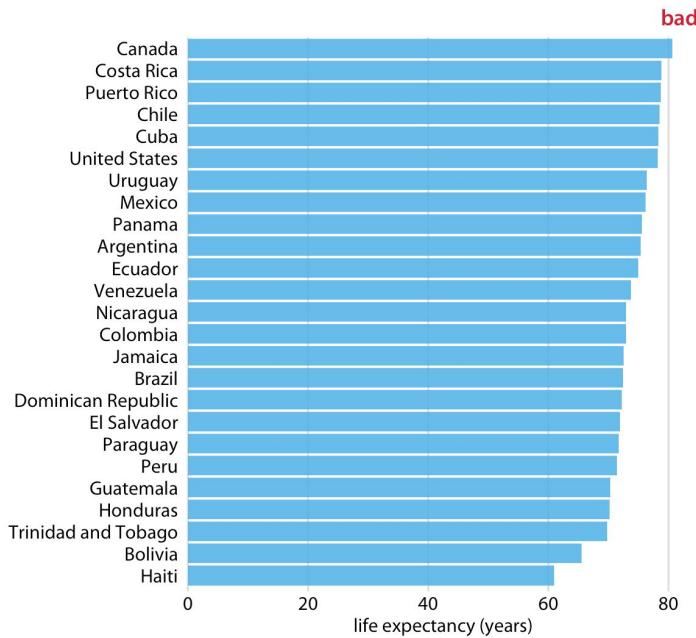
1.B Ordenar por magnitud: simplifica comparación entre las categorías y es visualmente más estético



```
sns.barplot(x='country_name',  
            y='value',  
            data=internet2016,  
            ax=ax[0],  
            ci=False)
```

```
sns.barplot(x='value',  
            y='country_name',  
            data=internet2016.\  
                sort_values(by='value',\.  
                           ascending=True),  
            ax=ax[1],  
            ci=False)
```

1.C Las barras deben contener el cero; seguir buenas prácticas no garantiza una visualización efectiva



- Se distorsionan las magnitudes si barras **NO** comienzan desde el origen
- En barras con niveles similares se dificulta distinguir las diferencias entre categorías
- ¿Qué podemos hacer?

Figure 6.12: Life expectancies of countries in the Americas, for the year 2007, shown as bars. This dataset is not suitable for being visualized with bars. The bars are too long and they draw attention away from the key feature of the data, the differences in life expectancy among the different countries. Data source: Gapminder project

1.C Liberarse del origen para ver las diferencias: utilizar puntos y ordenar las categorías

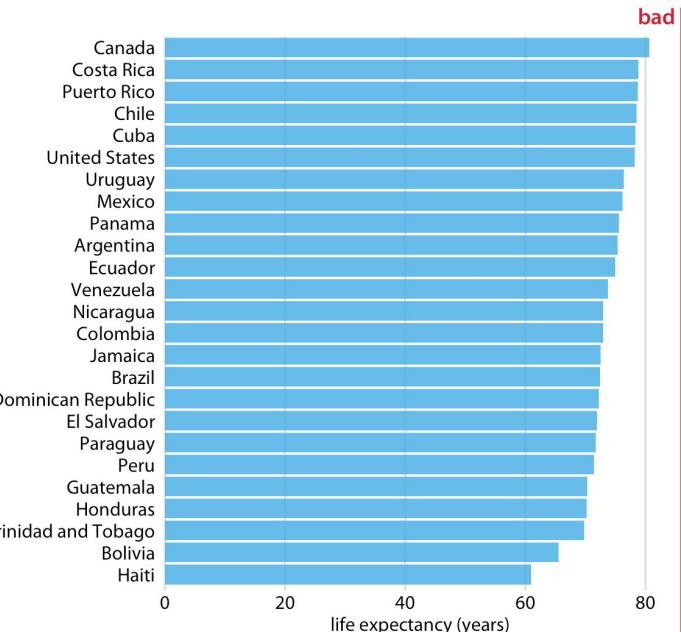


Figure 6.12: Life expectancies of countries in the Americas, for the year 2007, shown as bars. This dataset is not suitable for being visualized with bars. The bars are too long and they draw attention away from the key feature of the data, the differences in life expectancy among the different countries. Data source: Gapminder project

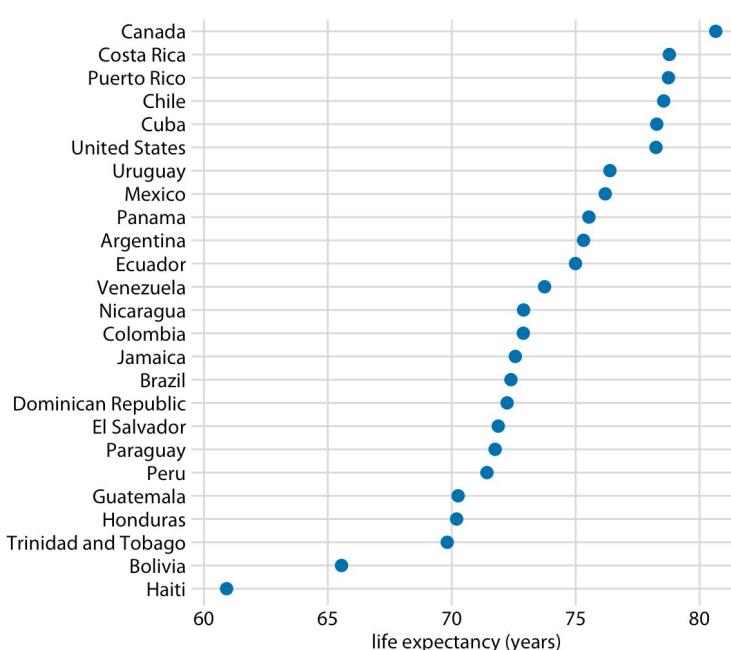
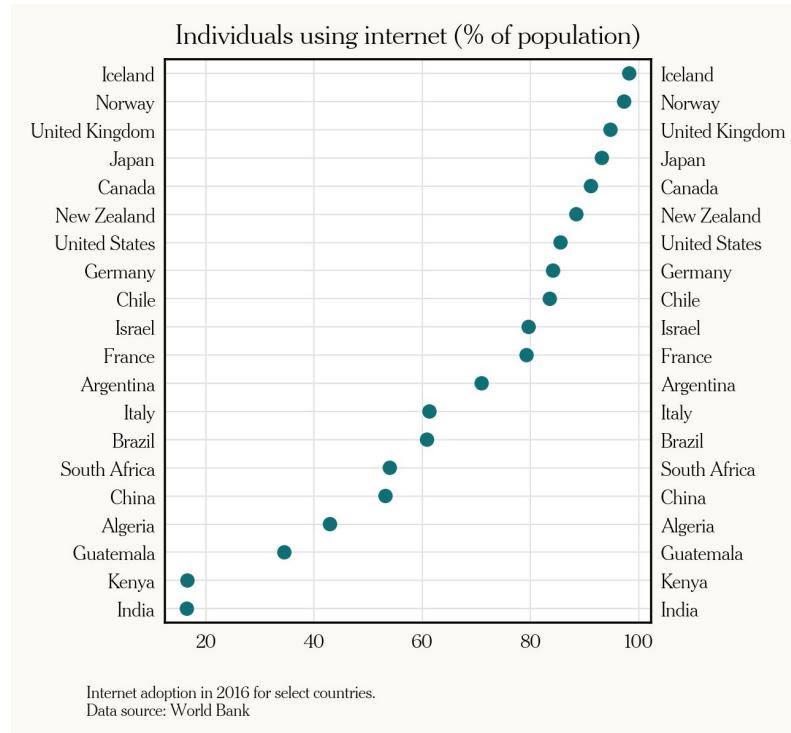


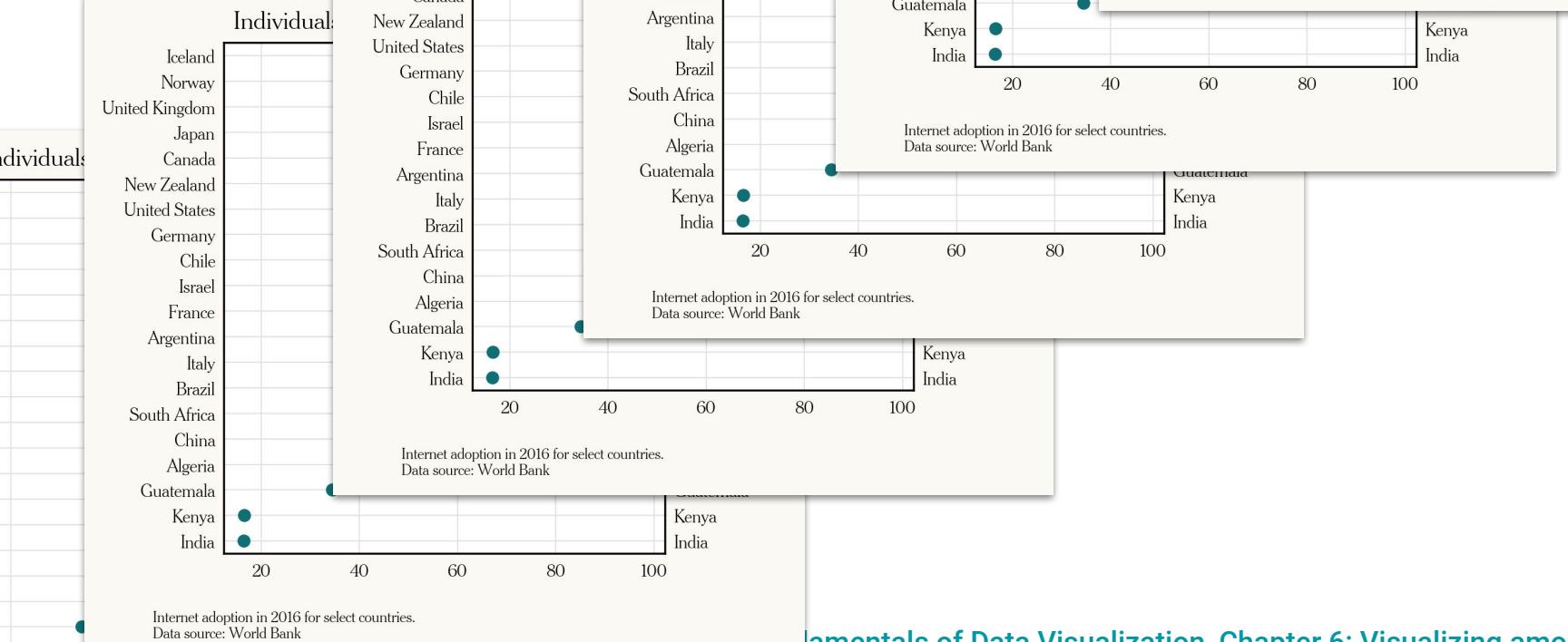
Figure 6.11: Life expectancies of countries in the Americas, for the year 2007. Data source: Gapminder project

1.C Liberarse del origen para ver las diferencias: utilizar puntos y ordenar las categorías



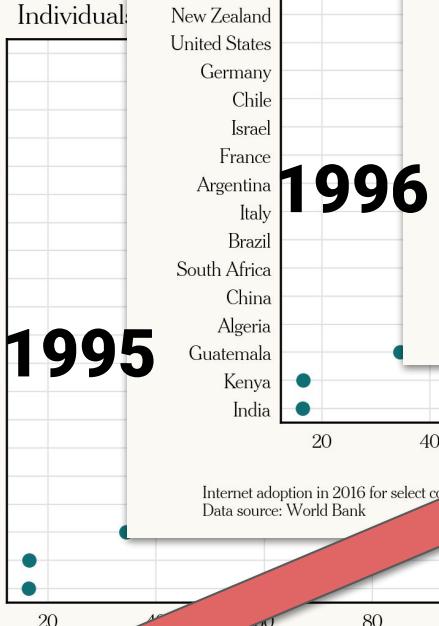
```
sns.pointplot(y='country_name',
                x='value',
                color='#107075',
                scale=.6,
                join=False,
                data=internet2016.\
                    sort_values(by='value',\
                               ascending=False)
```

1.D Múltiples barras foto a las + grandes



1.D Múltiples barras foto a las + grandes

1994
Individuals



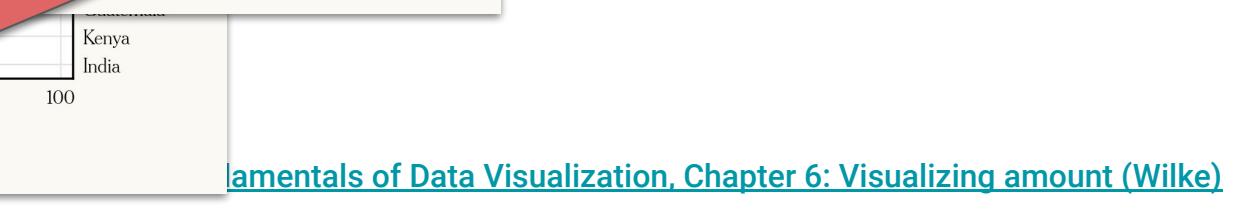
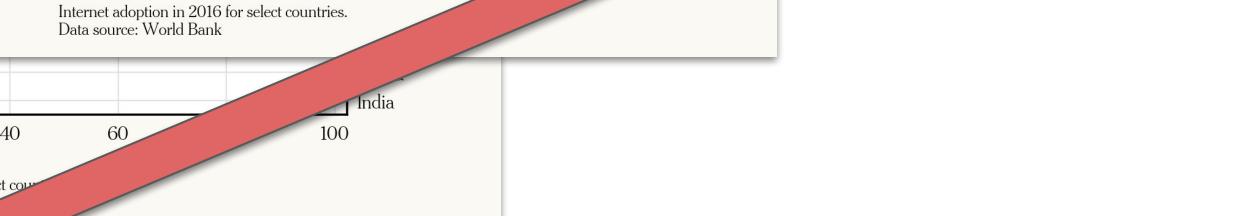
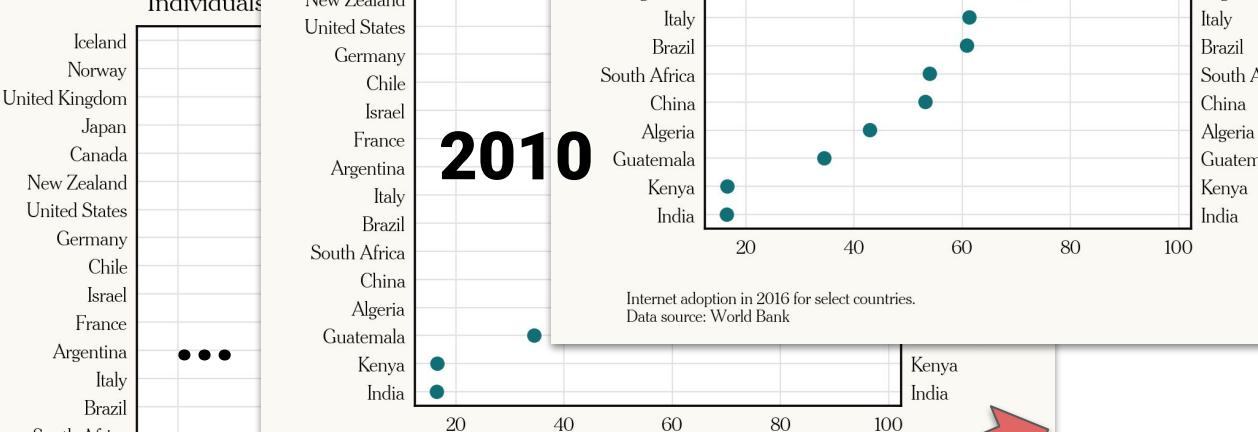
Internet adoption in 2016 for select countries.
Data source: World Bank



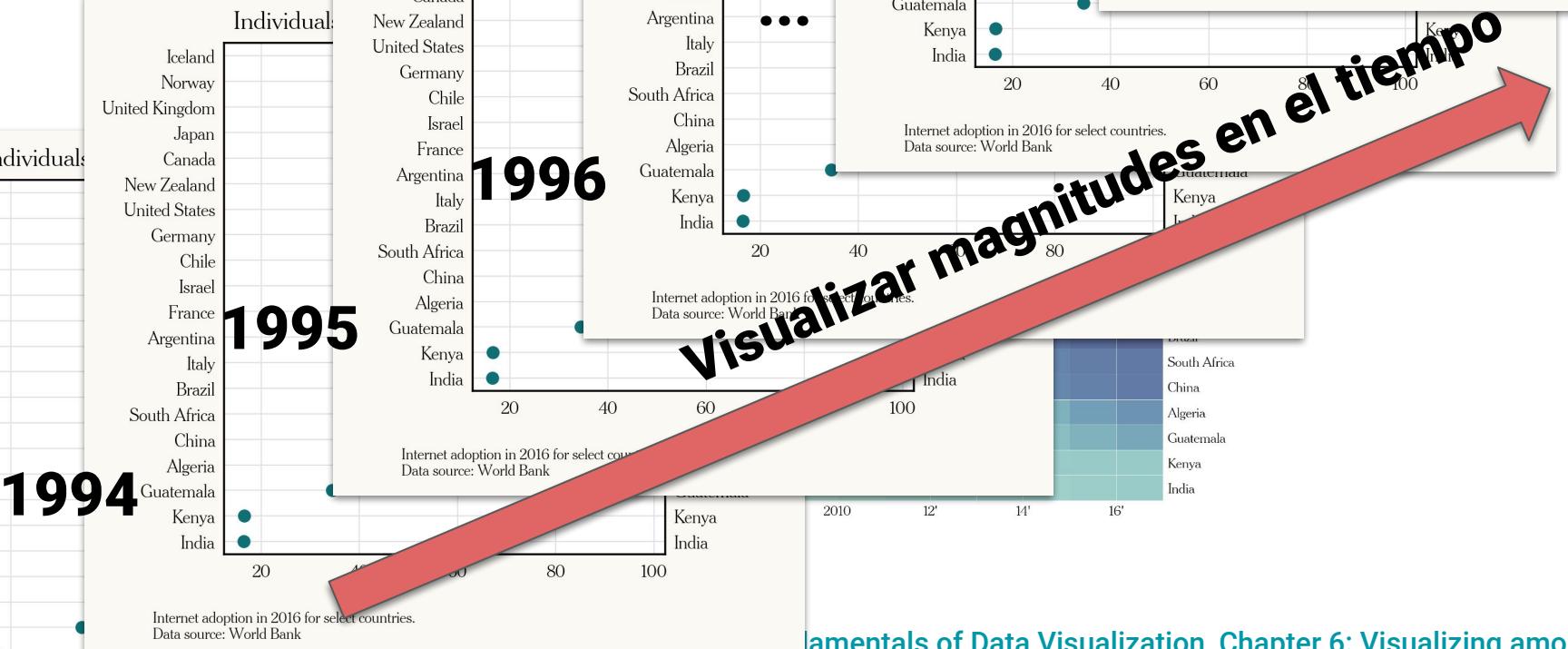
Internet adoption in 2016 for select countries.
Data source: World Bank



Internet adoption in 2016 for select countries.
Data source: World Bank



1.D Múltiples barras foto a las + grandes



1.D Múltiples barras en un heatmap: de la precisión de una foto a las tendencias generales

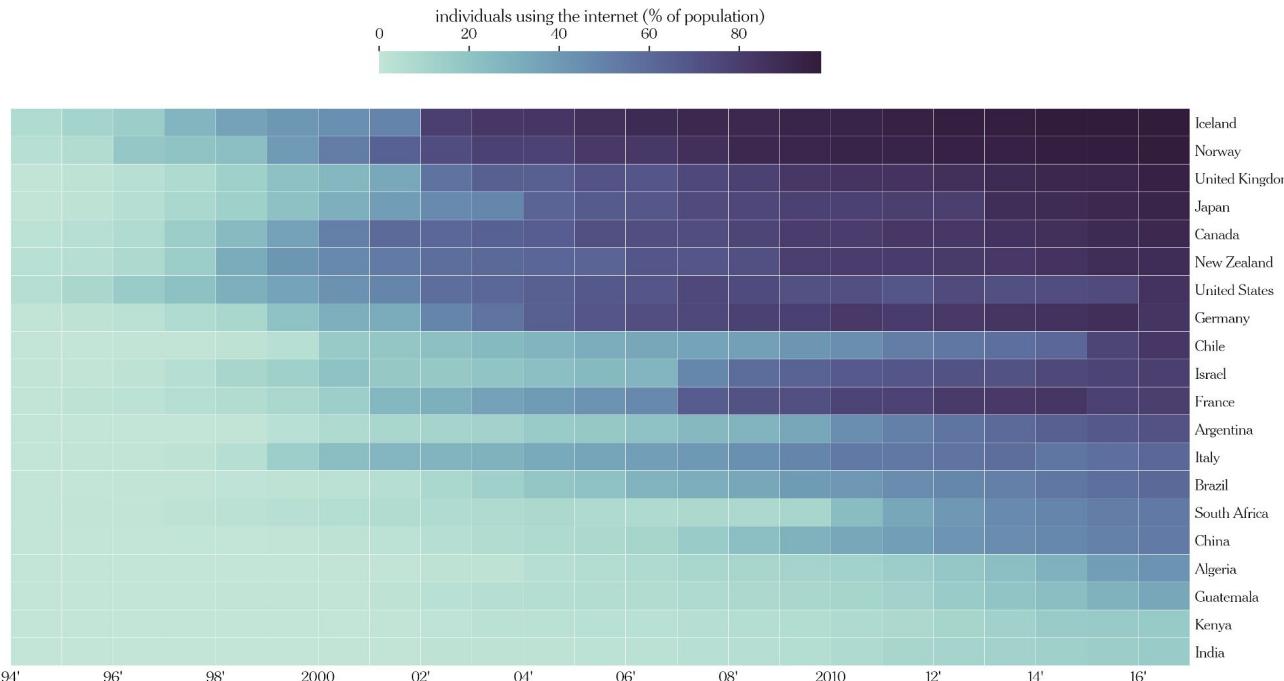


Figure 6.14: Internet adoption over time, for select countries. Color represents the percent of internet users for the respective country and year. Countries were ordered by percent internet users in 2016. Data source: World Bank

1.D Múltiples barras en un heatmap: orden de países en base al % de uso del 2016

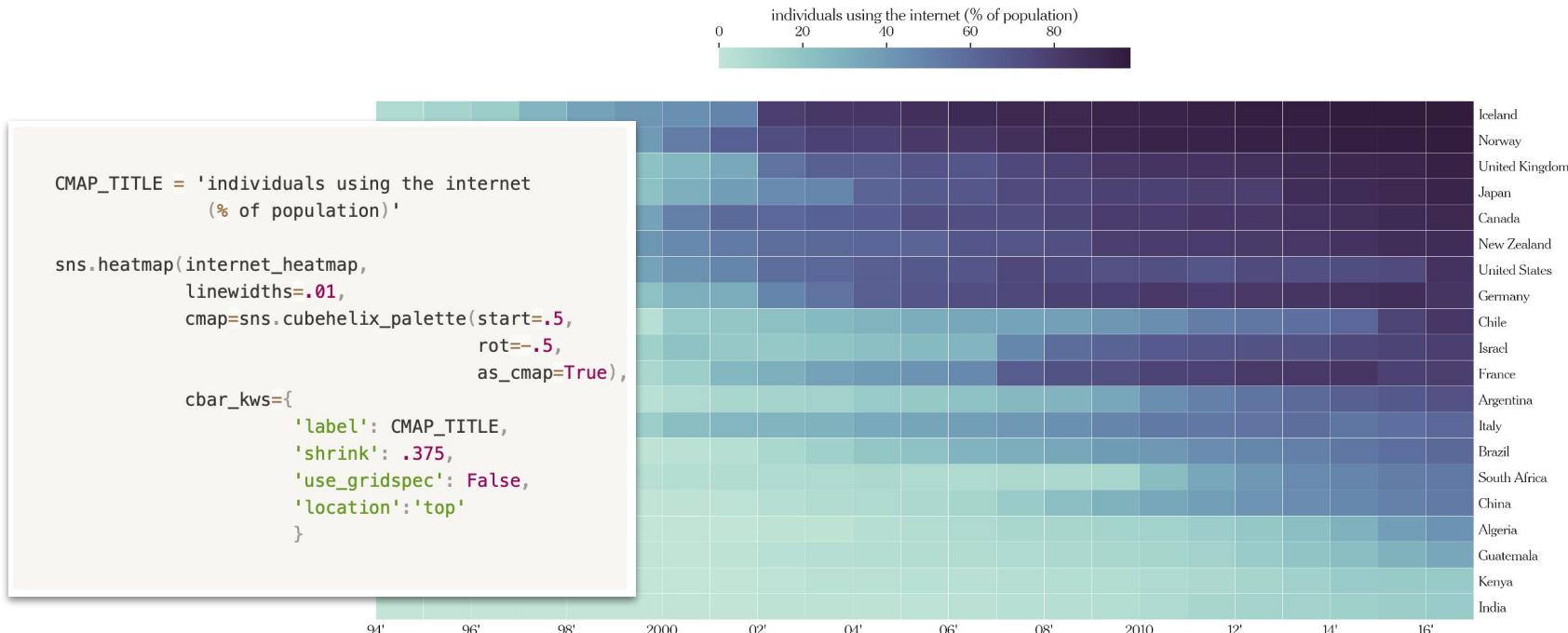


Figure 6.14: Internet adoption over time, for select countries. Color represents the percent of internet users for the respective country and year. Countries were ordered by percent internet users in 2016. Data source: World Bank

1.D El orden revela las tendencias: configuración de países respecto al primer año sobre 20% de uso de internet

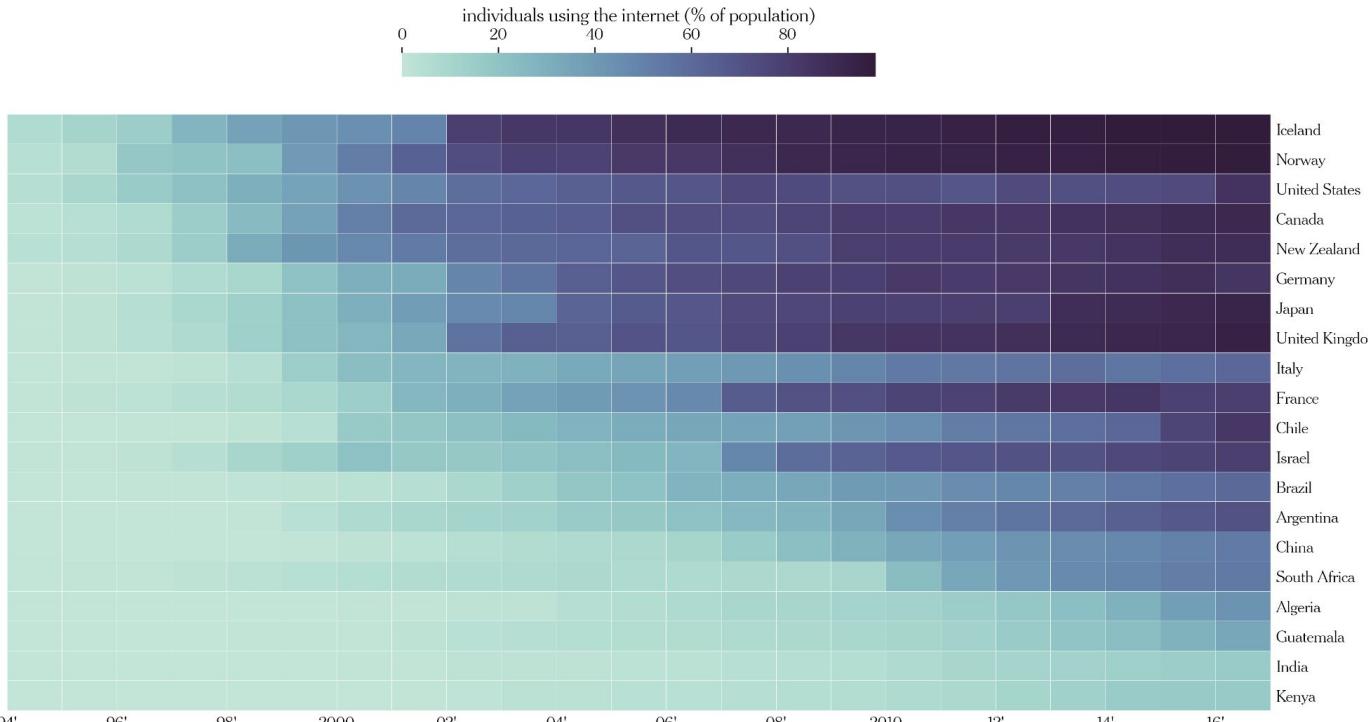


Figure 6.15: Internet adoption over time, for select countries. Countries were ordered by the year in which their internet usage first exceeded 20%.
Data source: World Bank

1.D El orden revela las tendencias: configuración de países respecto al primer año sobre 20% de uso de internet

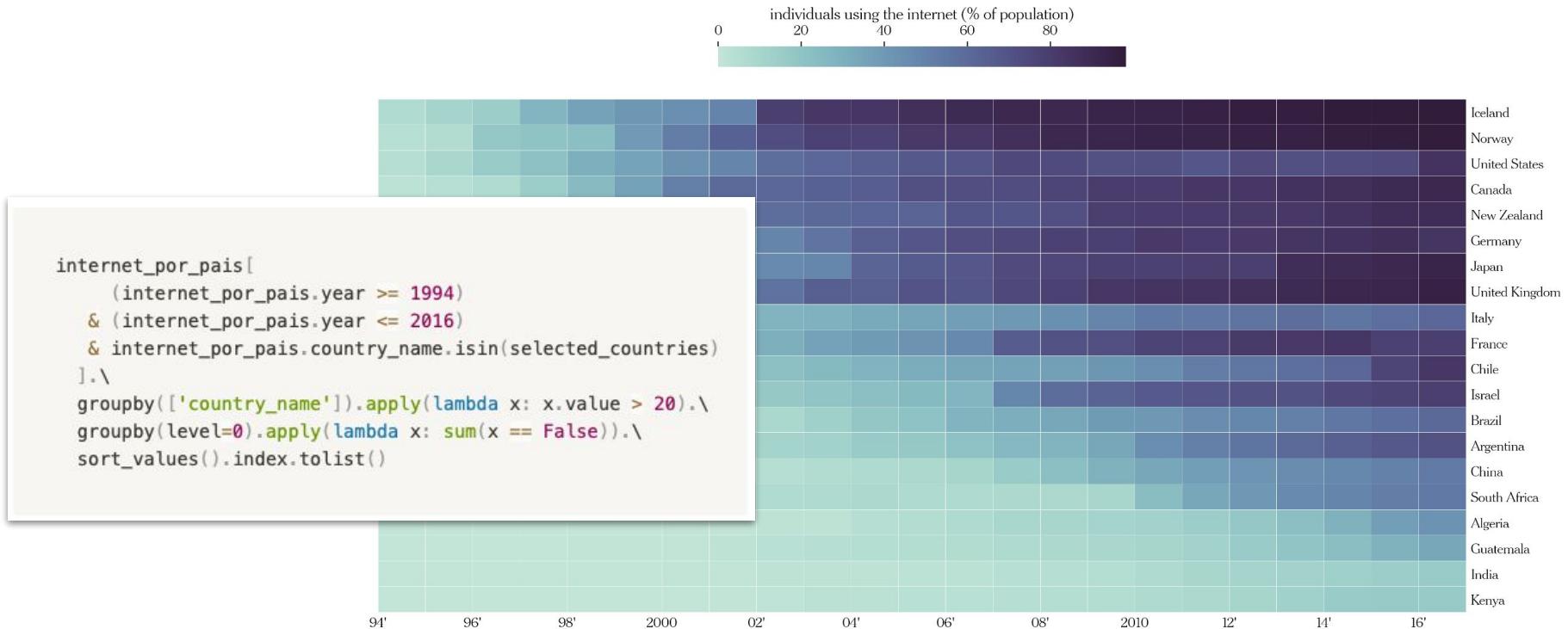


Figure 6.15: Internet adoption over time, for select countries. Countries were ordered by the year in which their internet usage first exceeded 20%.
Data source: World Bank

1.D El orden revela las tendencias

• configuración de países

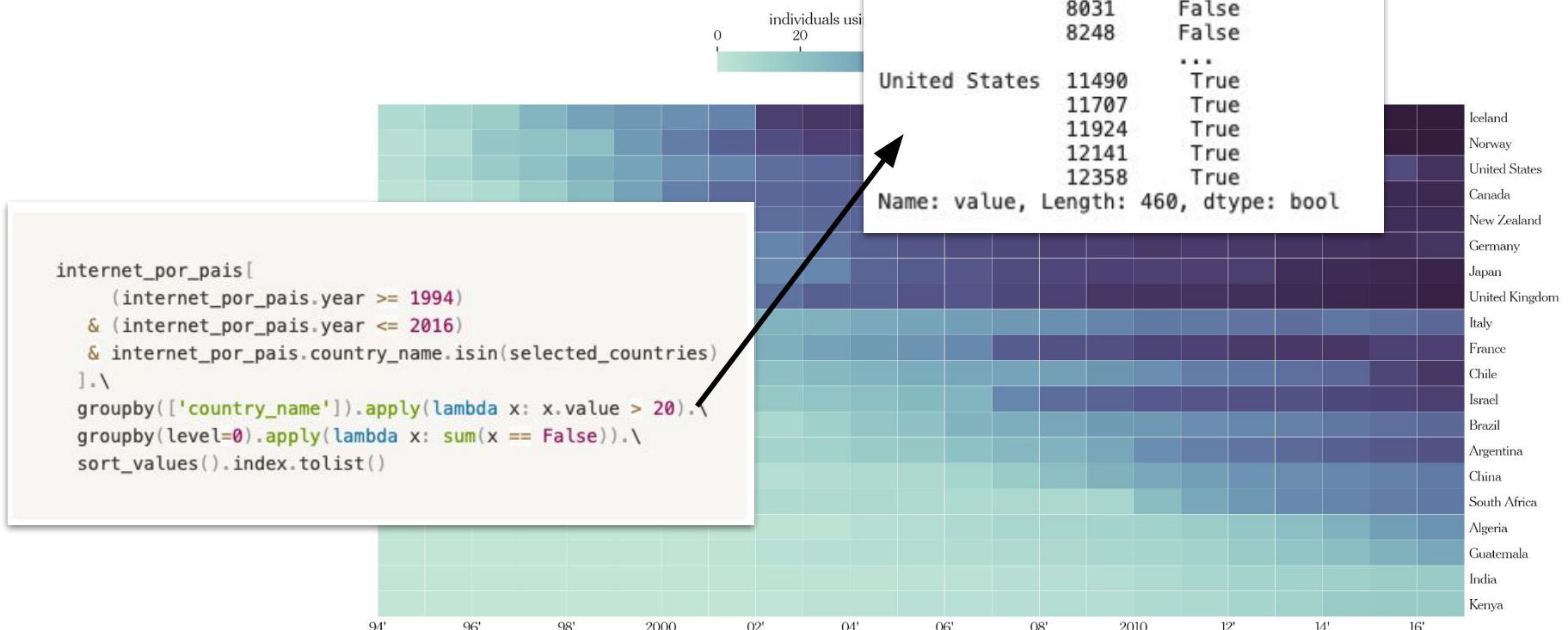


Figure 6.15: Internet adoption over time, for select countries. Countries were ordered by the year in which their internet usage first exceeded 20%.
Data source: World Bank

1.D El orden revela las tendencias

• configuración de países

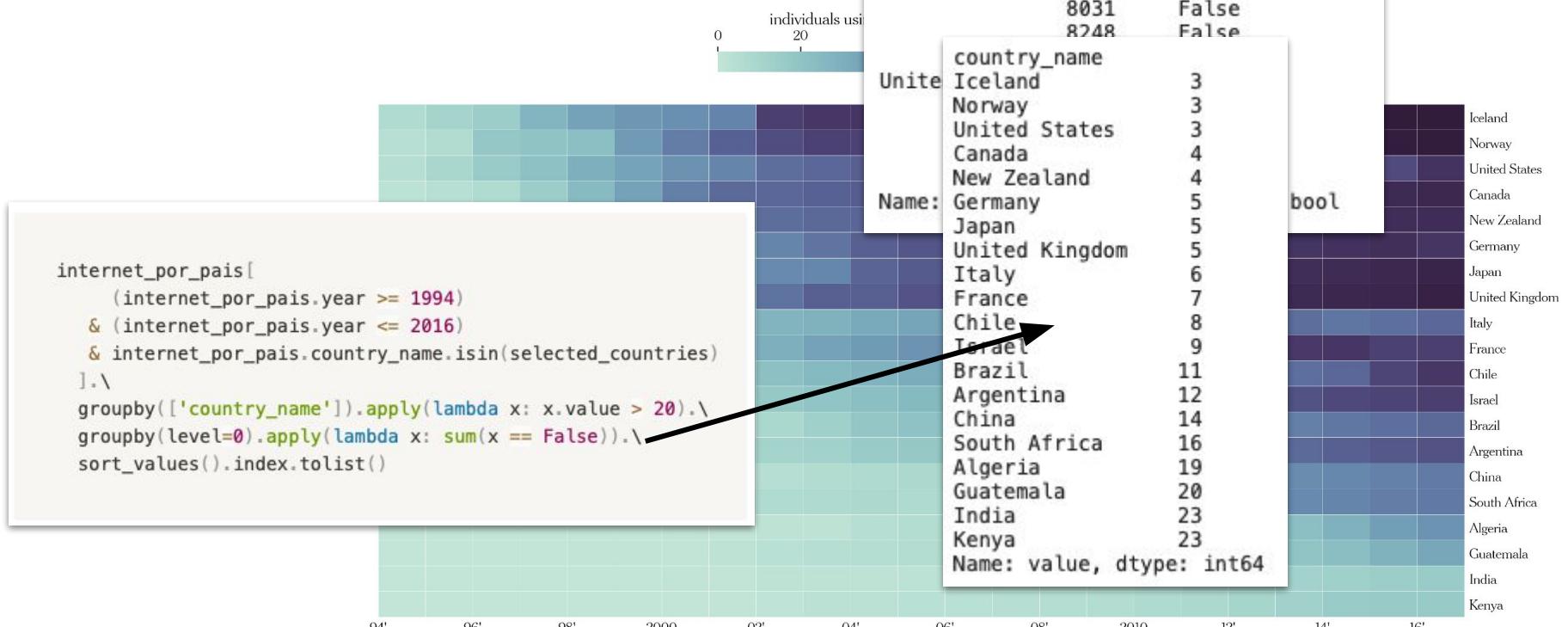


Figure 6.15: Internet adoption over time, for select countries. Countries were ordered by the year in which their internet usage first exceeded 20%.
Data source: World Bank

1.D El orden revela las tendencias

• configuración de países respecto al primer año sobre 20% de uso

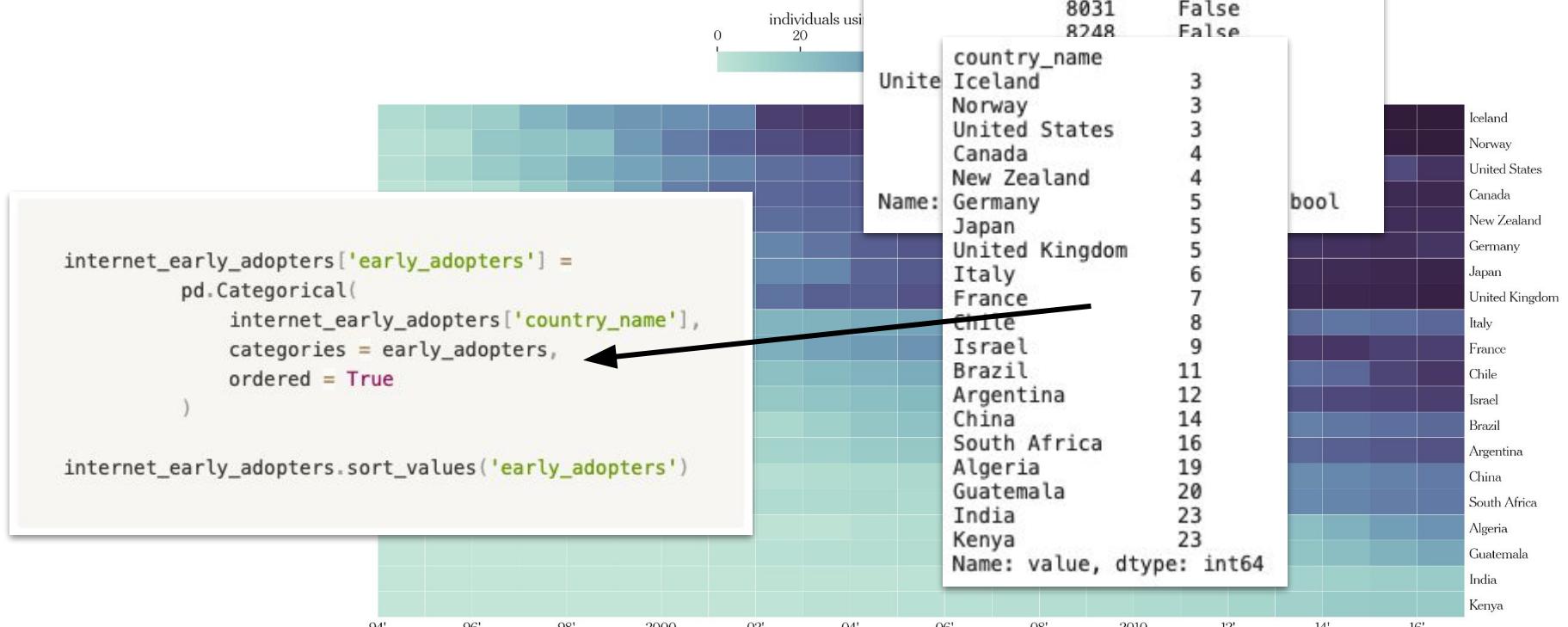
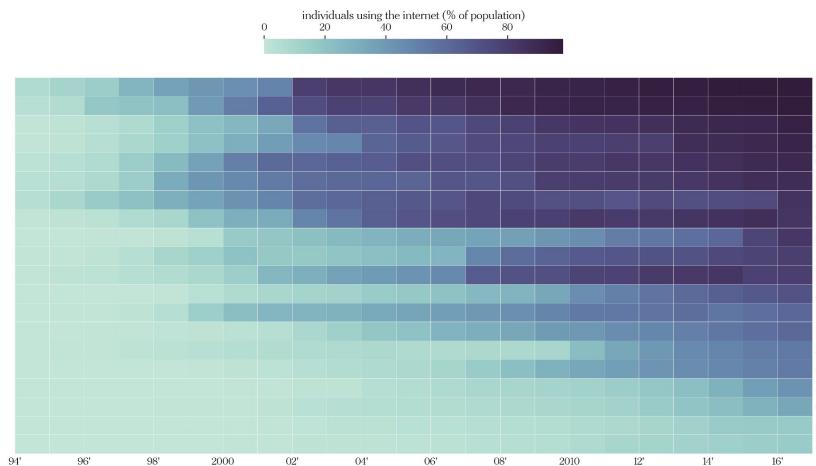
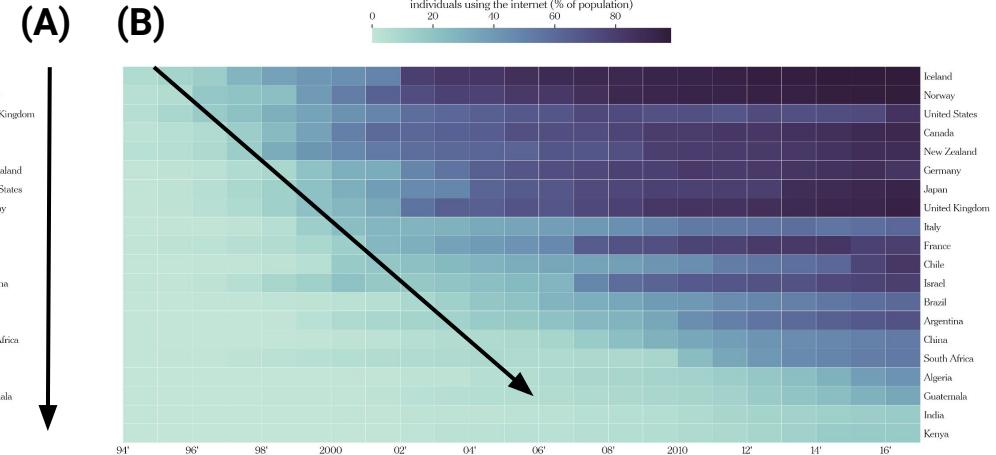


Figure 6.15: Internet adoption over time, for select countries. Countries were ordered by the year in which their internet usage first exceeded 20%.
Data source: World Bank

1.D Múltiples barras en un heatmap: comparación entre diferentes configuraciones de orden



(A) Orden de los países en base al % uso de internet año 2016

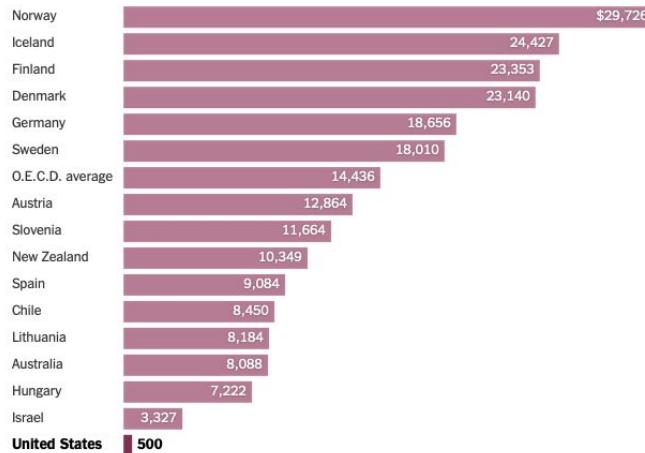


(B) Orden de los países en base al primer año con 20% o más de adopción de internet

1.E Usar anotaciones: desde entregar mayor contexto a eliminar un eje

How Much Governments Spend on Child Care for Toddlers

Annual public spending per child on early childhood care.



Anotaciones con explicaciones, detalles adicionales, y fuente de los datos

In the United States, most of the spending is on families living in poverty. Some O.E.C.D. countries are not included because data on their spending was unavailable.

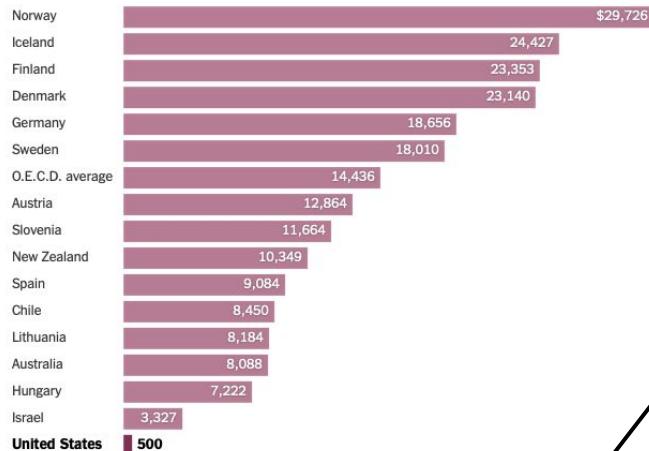
By The New York Times | Source: Organization for Economic Cooperation and Development and Elizabeth Davis and Aaron Sojourner for the Hamilton Project.

[The New York Times: How Other Nations Pay for Child Care. The U.S. Is an Outlier](#)

1.E Usar anotaciones: desde entregar mayor contexto a eliminar un eje

How Much Governments Spend on Child Care for Toddlers

Annual public spending per child on early childhood care.



Anotaciones con explicaciones, detalles adicionales, y fuente de los datos

In the United States, most of the spending is on families living in poverty. Some O.E.C.D. countries are not included because data on their spending was unavailable.

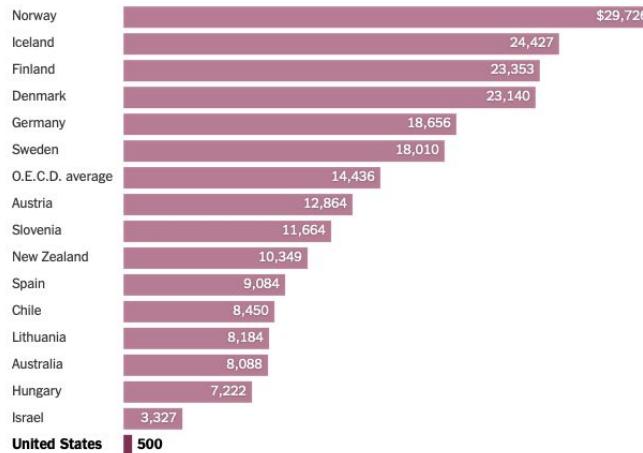
By The New York Times | Source: Organization for Economic Cooperation and Development and Elizabeth Davis and Aaron Sojourner for the Hamilton Project.

[The New York Times: How Other Nations Pay for Child Care. The U.S. Is an Outlier](#)

1.E Usar anotaciones: desde entregar mayor contexto a eliminar un eje

How Much Governments Spend on Child Care for Toddlers

Annual public spending per child on early childhood care.



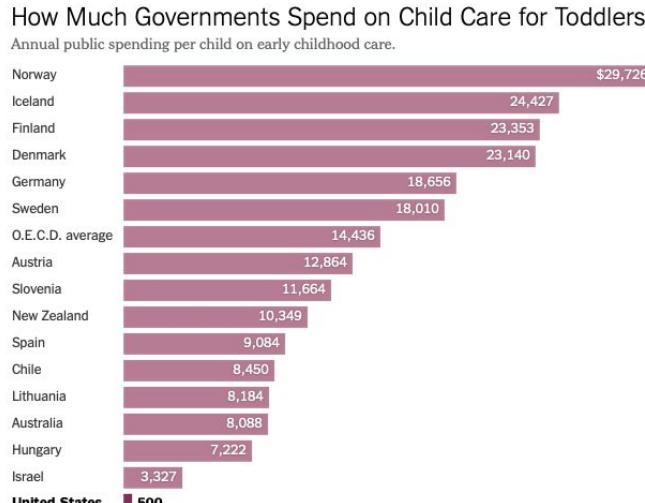
In the United States, most of the spending is on families living in poverty. Some O.E.C.D. countries are not included because data on their spending was unavailable.

By The New York Times | Source: Organization for Economic Cooperation and Development and Elizabeth Davis and Aaron Sojourner for the Hamilton Project.

Anotaciones con explicaciones, detalles adicionales, y fuente de los datos

El eje x se puede “apagar” si se agregan anotaciones con las magnitudes de cada barra

1.E Usar anotaciones: desde entregar mayor contexto a eliminar un eje



In the United States, most of the spending is on families living in poverty. Some O.E.C.D. countries are not included because data on their spending was unavailable.

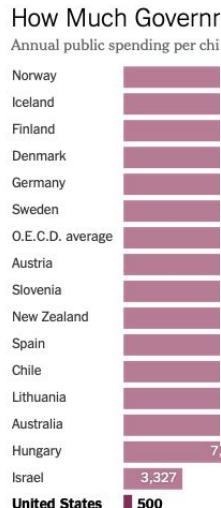
By The New York Times | Source: Organization for Economic Cooperation and Development and Elizabeth Davis and Aaron Sojourner for the Hamilton Project.

Anotaciones con explicaciones, detalles adicionales, y fuente de los datos

El eje x se puede “apagar” si se agregan anotaciones con las magnitudes de cada barra

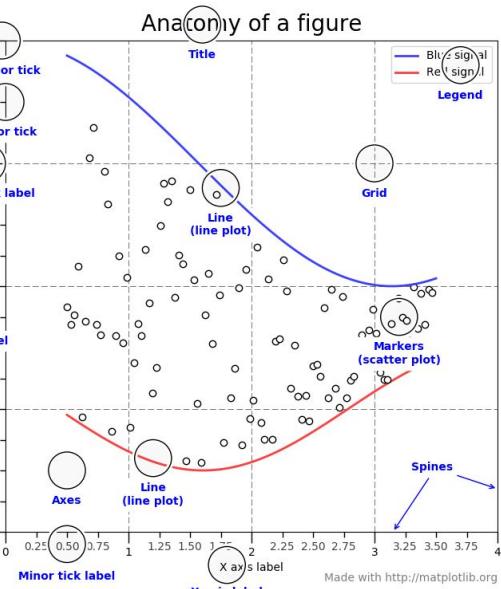
```
ax.spines['bottom'].set_visible(False)  
ax.axes.get_xaxis().set_visible(False)
```

1.E Usar annotations para eliminar un outlier



In the United States, most of the spending is on families living in poverty. Some O.E.C.D. countries are not included because data on their spending was unavailable.

By The New York Times | Source: Organization for Economic Cooperation and Development and Elizabeth Davis and Aaron Sojourner for the Hamilton Project.



[The New York Times: How Other Nations Pay for Child Care. The U.S. Is an Outlier](#)

Anotaciones con explicaciones, detalles adicionales, y fuente de los datos

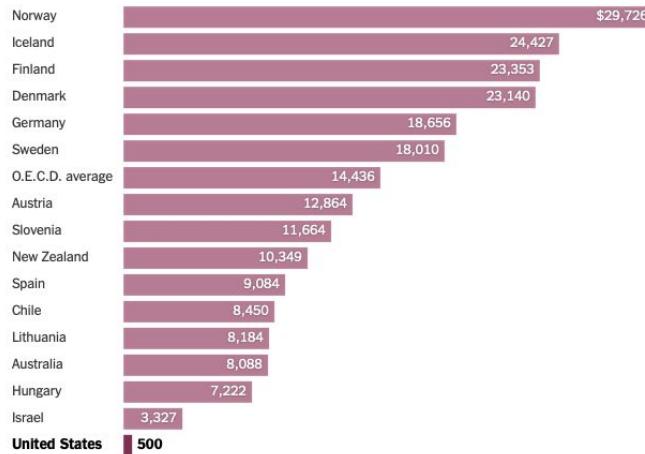
El eje x se puede “apagar” si se agregan anotaciones con las magnitudes de cada barra

```
ax.spines['bottom'].set_visible(False)  
ax.axes.get_xaxis().set_visible(False)
```

1.E Usar anotaciones: desde entregar mayor contexto a eliminar un eje

How Much Governments Spend on Child Care for Toddlers

Annual public spending per child on early childhood care.



In the United States, most of the spending is on families living in poverty. Some O.E.C.D. countries are not included because data on their spending was unavailable.

By The New York Times | Source: Organization for Economic Cooperation and Development and Elizabeth Davis and Aaron Sojourner for the Hamilton Project.

Anotaciones con explicaciones, detalles adicionales, y fuente de los datos

El eje x se puede “apagar” si se agregan anotaciones con las magnitudes de cada barra

Usar las mismas magnitudes menos un OFFSET para el posicionamiento (e.g. 29727 - 1000)

#2

Comparar distribuciones

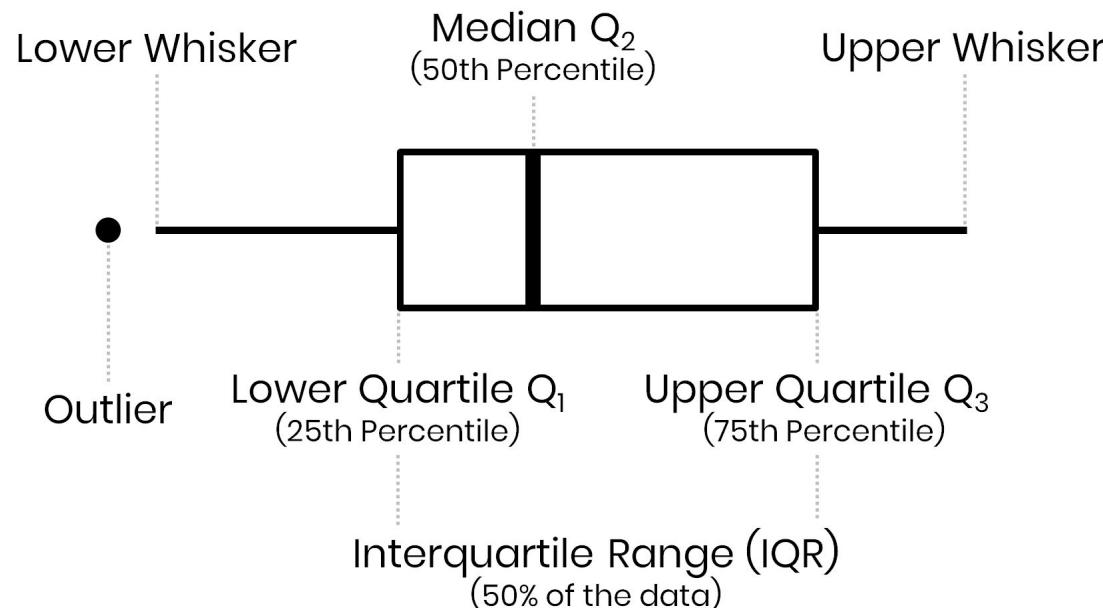


Más de una alternativa y más de una
audiencia

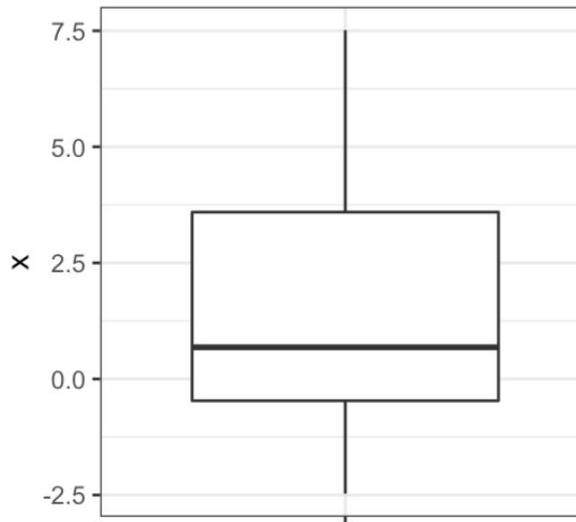
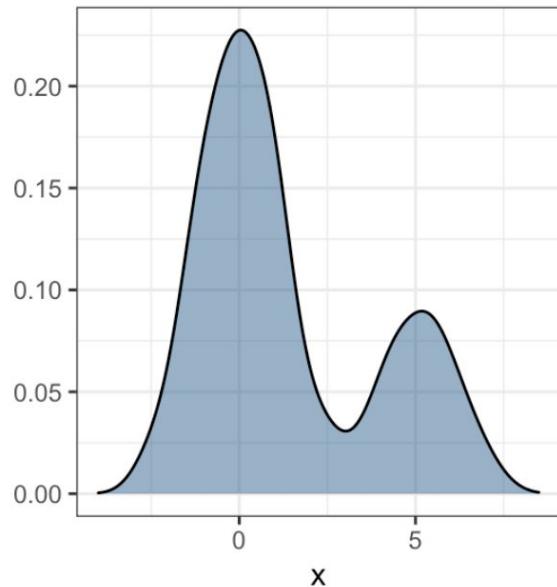
Visualización por capas

Anotaciones y anotaciones

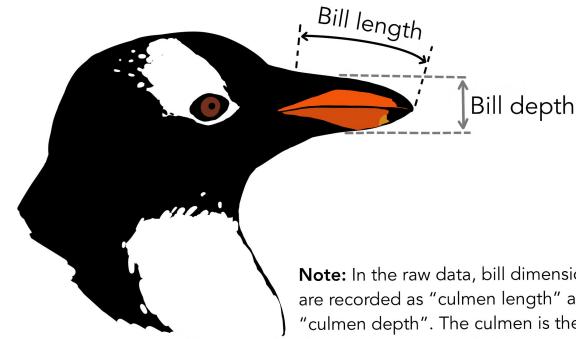
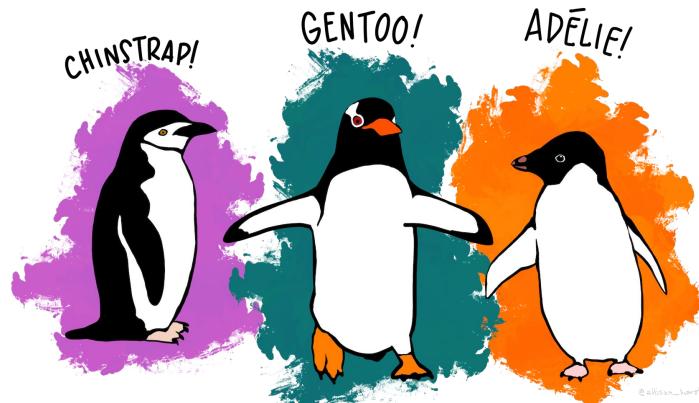
2.A Anatomía del gráfico de cajas y bigotes (aka **boxplot**)



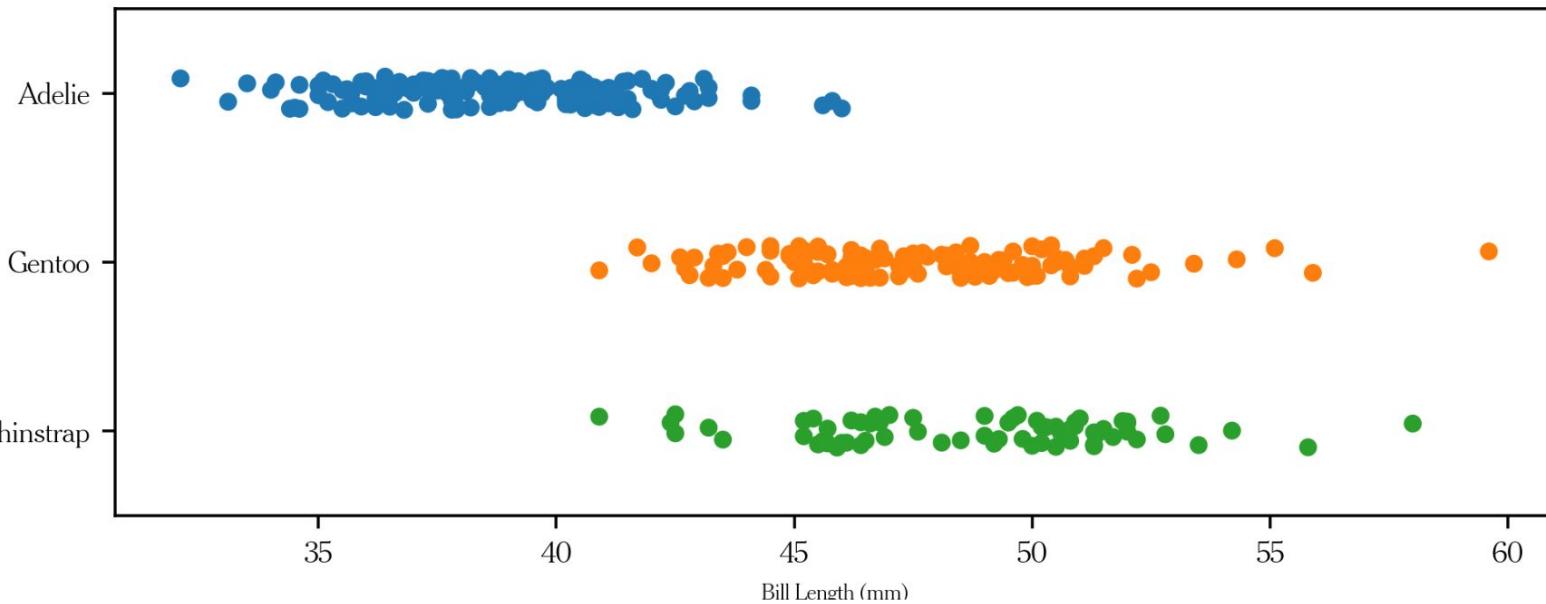
2.B Desventaja del boxplot: no permite identificar distribuciones bi-modales o multimodales



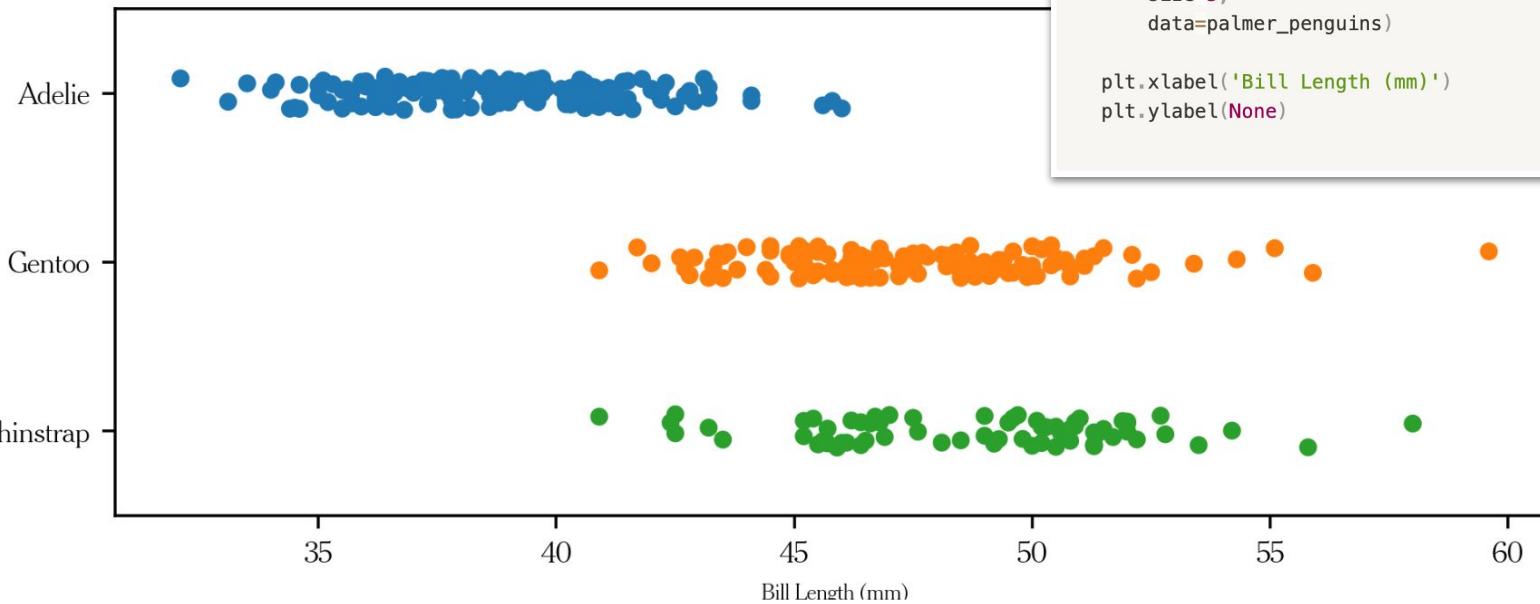
2.C Alternativa para comparar grupos: visualización por capas y anotaciones



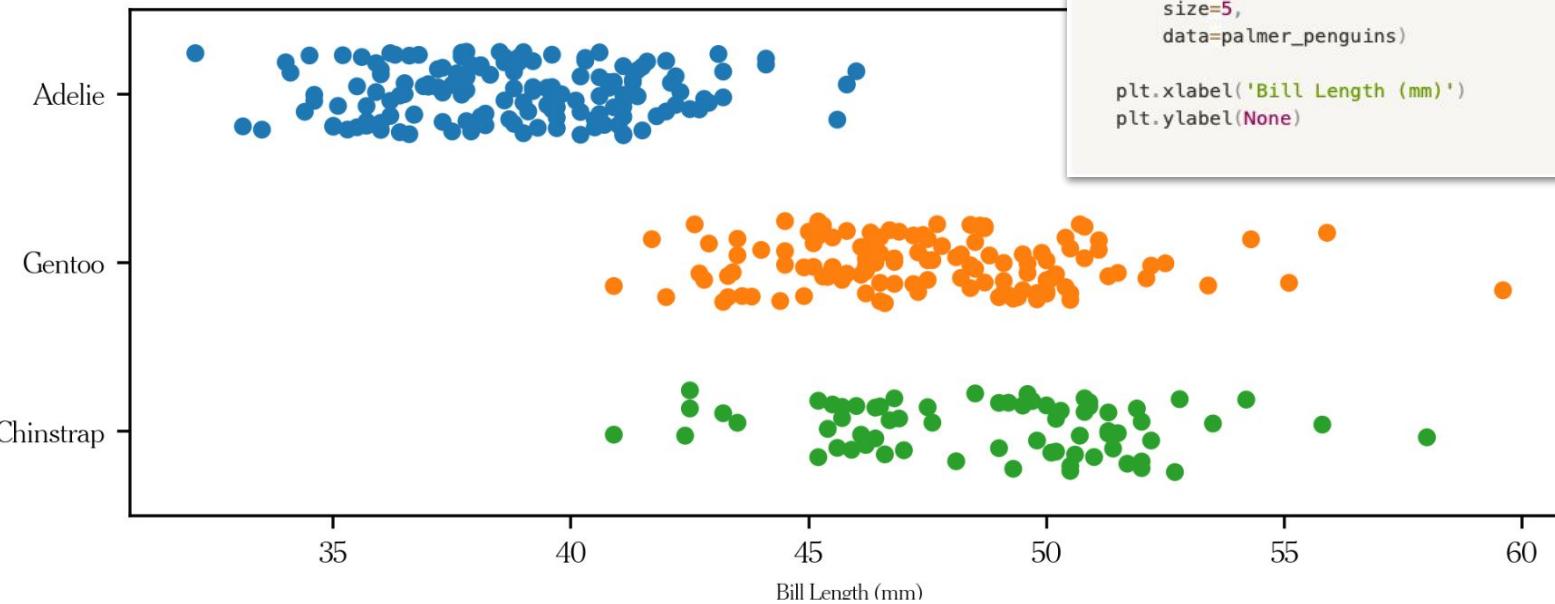
2.C Alternativa para comparar grupos: strip plot



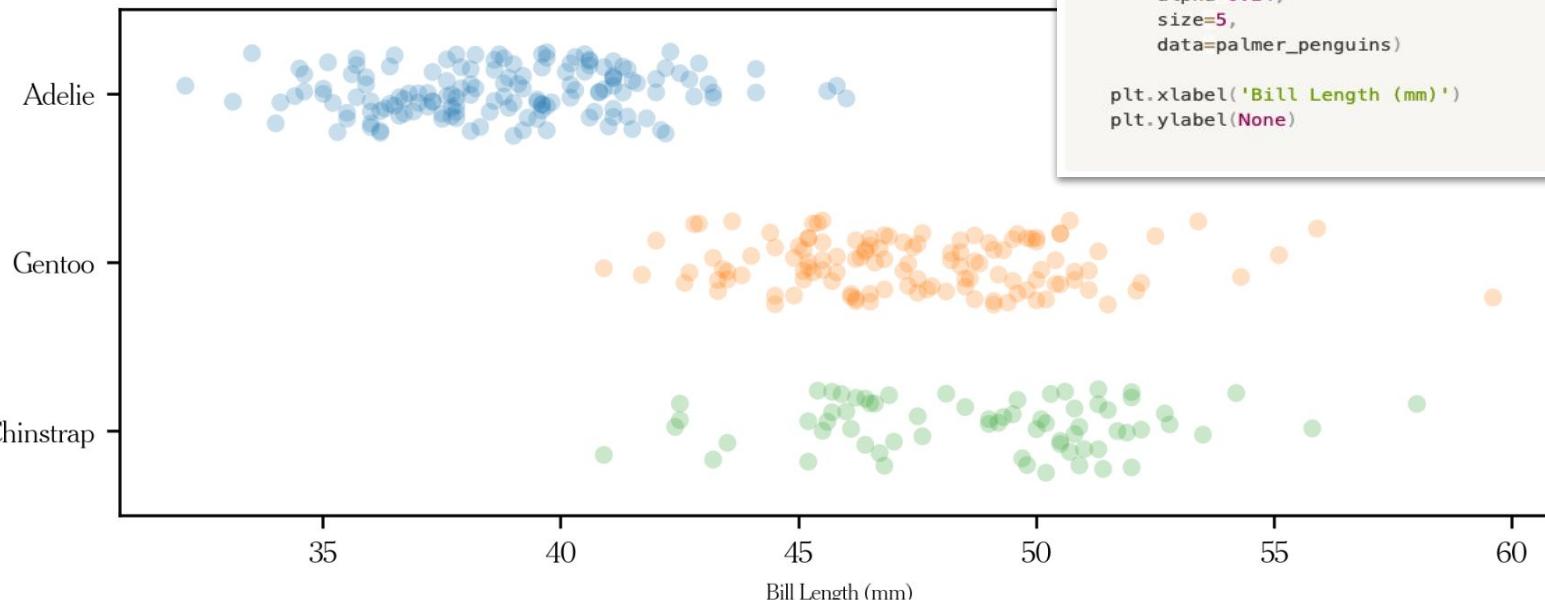
2.C Alternativa para comparar grupos: strip plot parámetros por defecto



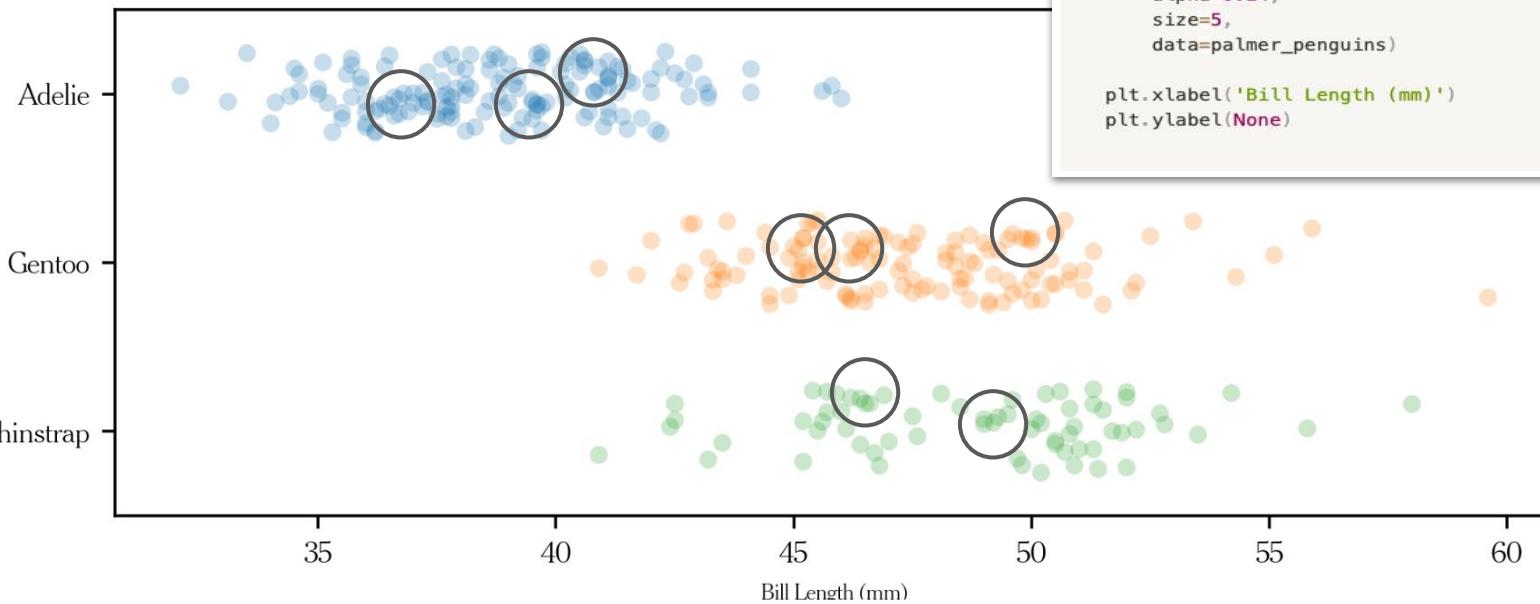
2.C Parámetro jitter: permite controlar el nivel de perturbación en los puntos (i.e. separación)



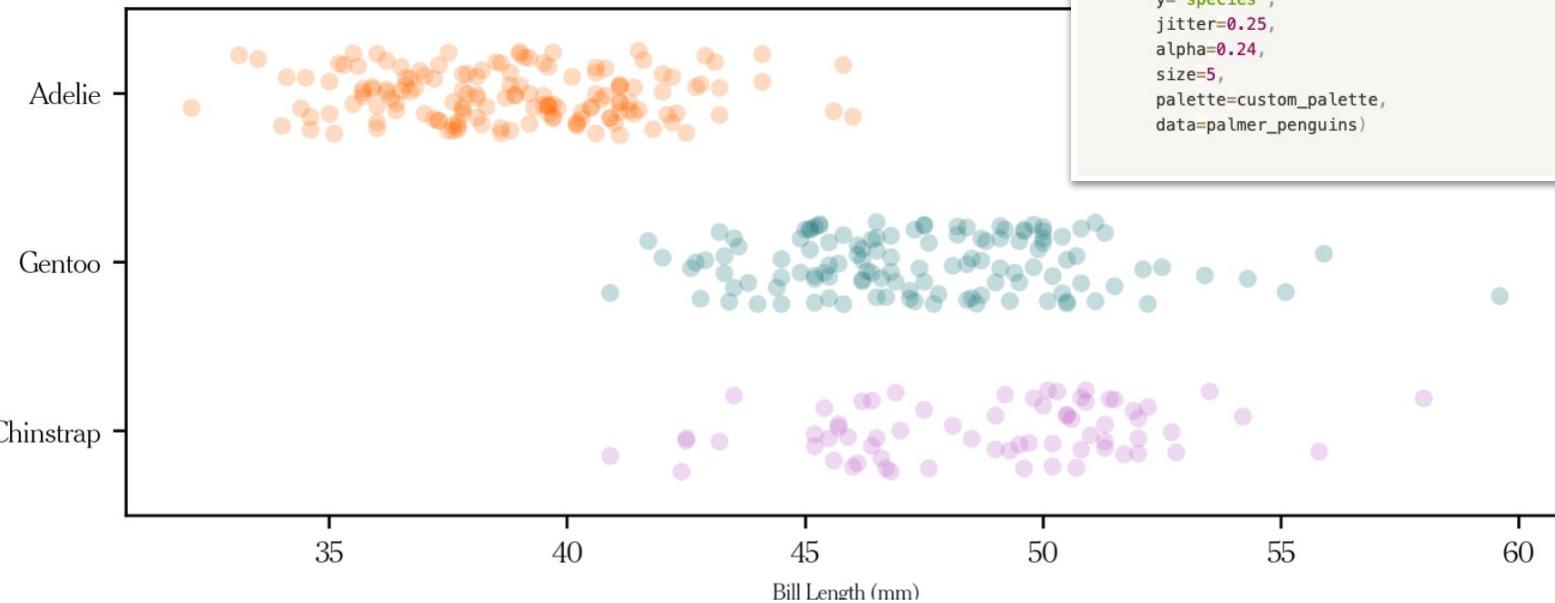
2.C Parámetro alpha: agregar un canal de transparencia a los colores (0: transparencia total)



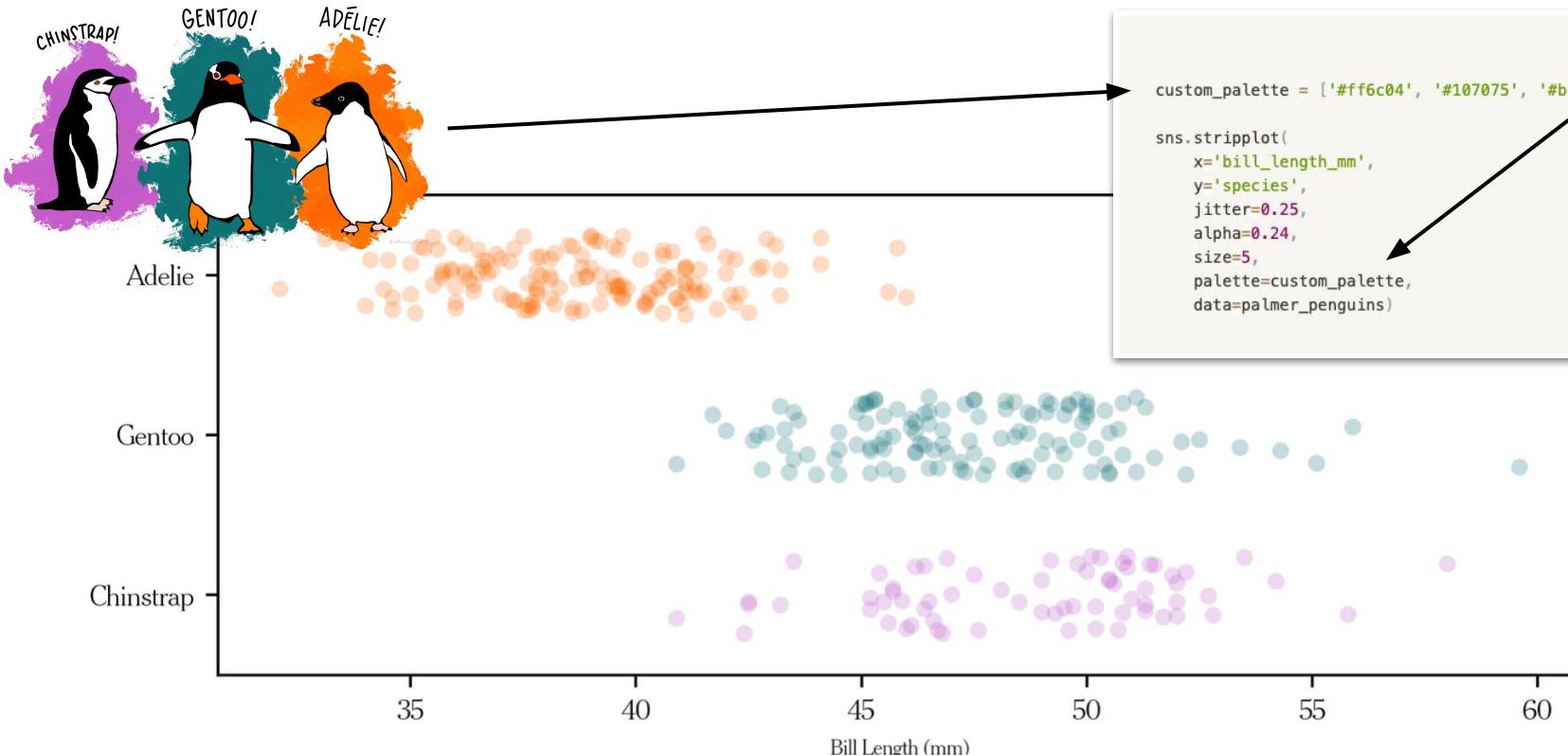
2.C Superposición: transparencia nos permite identificar mejor las concentraciones de puntos



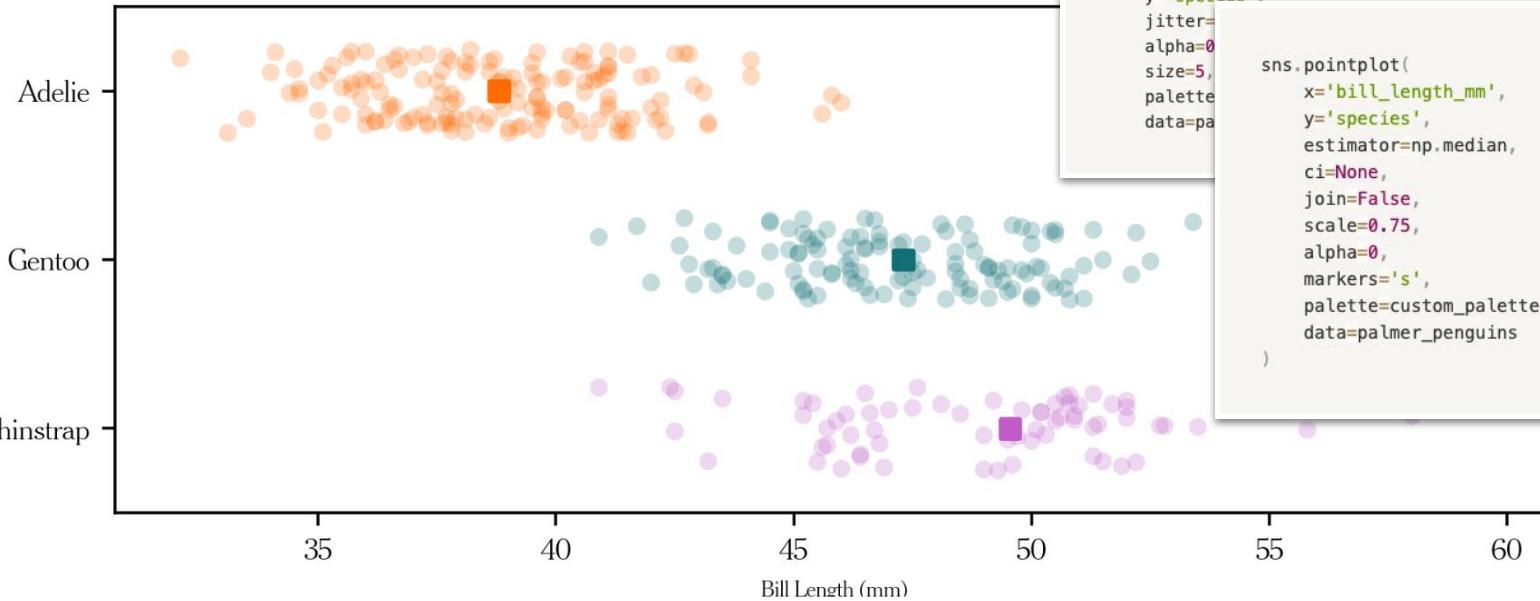
2.C Utilizar paleta de colores personalizada



2.C Utilizar paleta de colores personalizada



2.C Agregar otra capa de visualización: el punto (cuadrado) medio de cada distribución

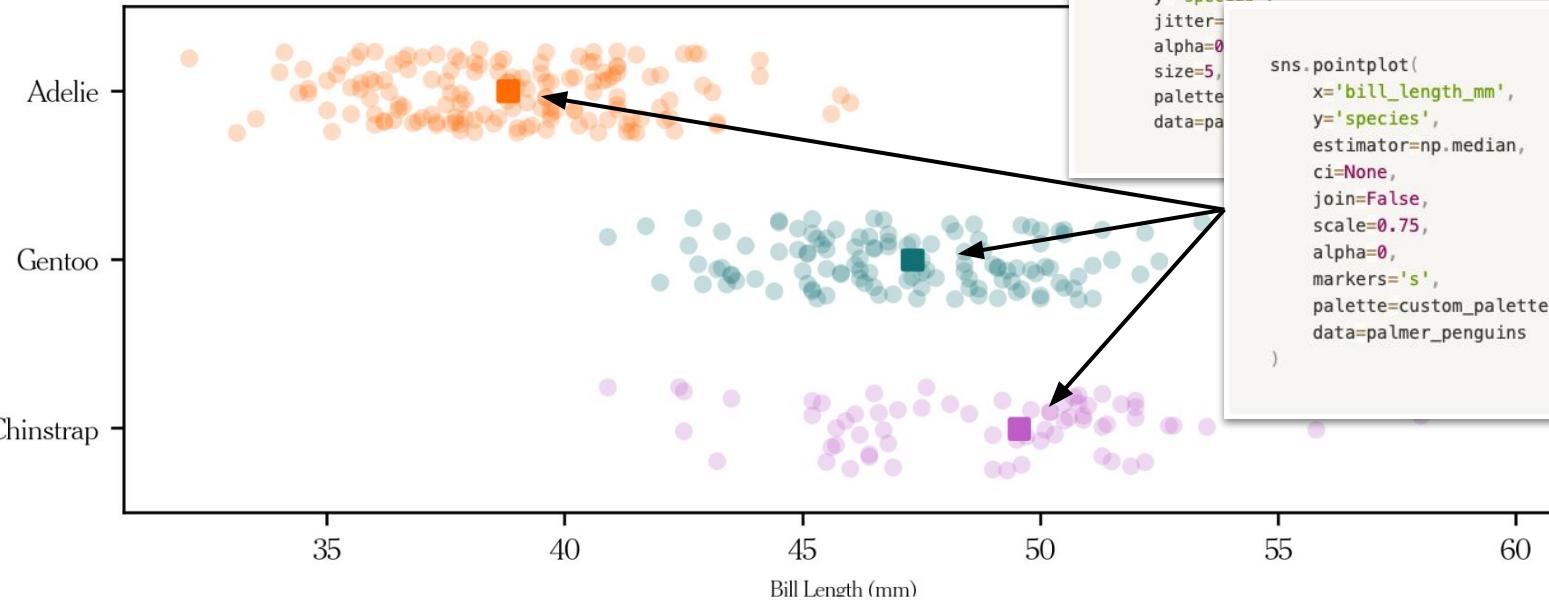


```
(1)
custom_palette = ['#ff6c04', '#107075', '#bf5cc9']

sns.stripplot(
    x='bill_length_mm',
    y='species',
    jitter=True,
    alpha=0.5,
    size=5,
    palette=custom_palette,
    data=pa
```

```
(2)
sns.pointplot(
    x='bill_length_mm',
    y='species',
    estimator=np.median,
    ci=None,
    join=False,
    scale=0.75,
    alpha=0,
    markers='s',
    palette=custom_palette,
    data=palmer_penguins
)
```

2.C Agregar otra capa de visualización: sns.pointplot por debajo agrega la información



```
custom_palette = ['#ff6c04', '#107075', '#bf5cc9']
```

```
sns.stripplot(  
    x='bill_length_mm',  
    y='species',  
    jitter=  
    alpha=0,  
    size=5,  
    palette=  
    data=pa
```

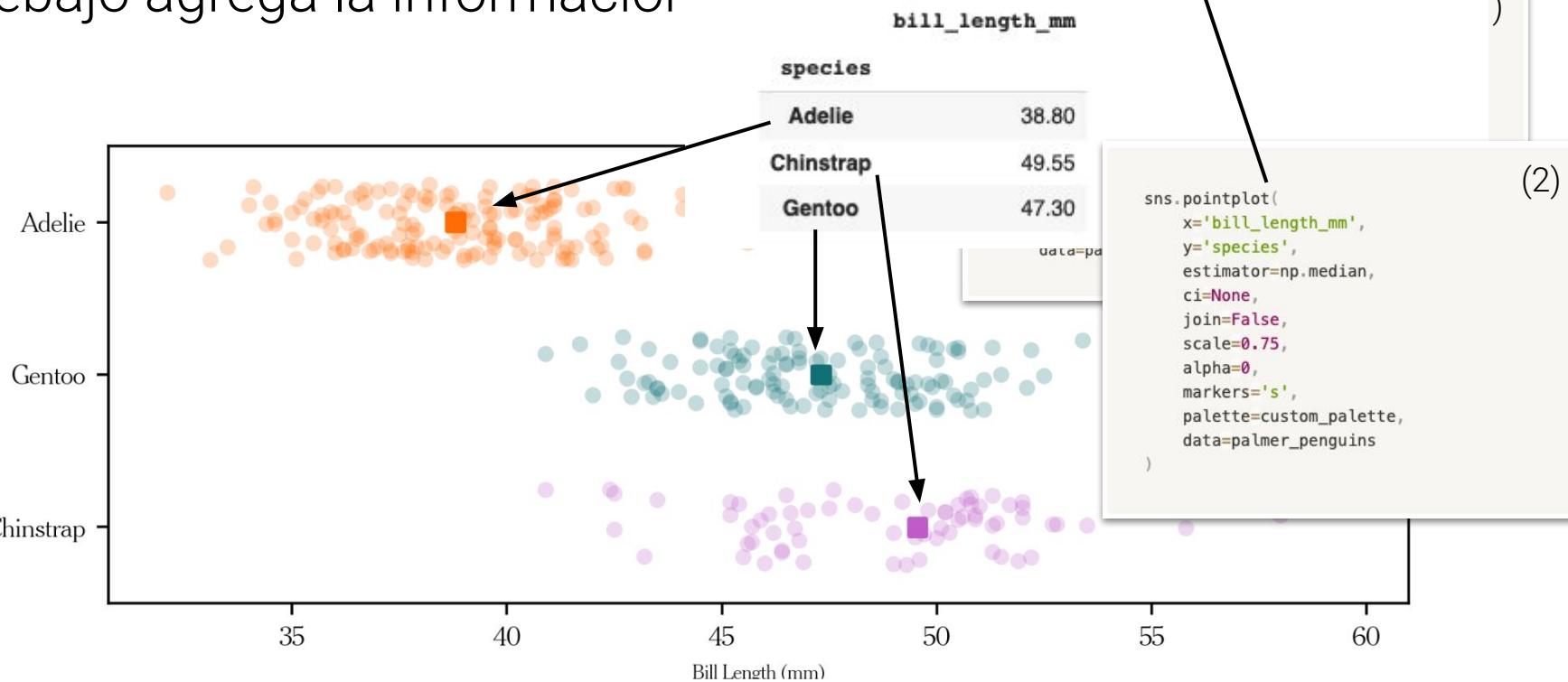
```
)  
  
sns.pointplot(  
    x='bill_length_mm',  
    y='species',  
    estimator=np.median,  
    ci=None,  
    join=False,  
    scale=0.75,  
    alpha=0,  
    markers='s',  
    palette=custom_palette,  
    data=palmer_penguins  
)
```

(1)

(2)

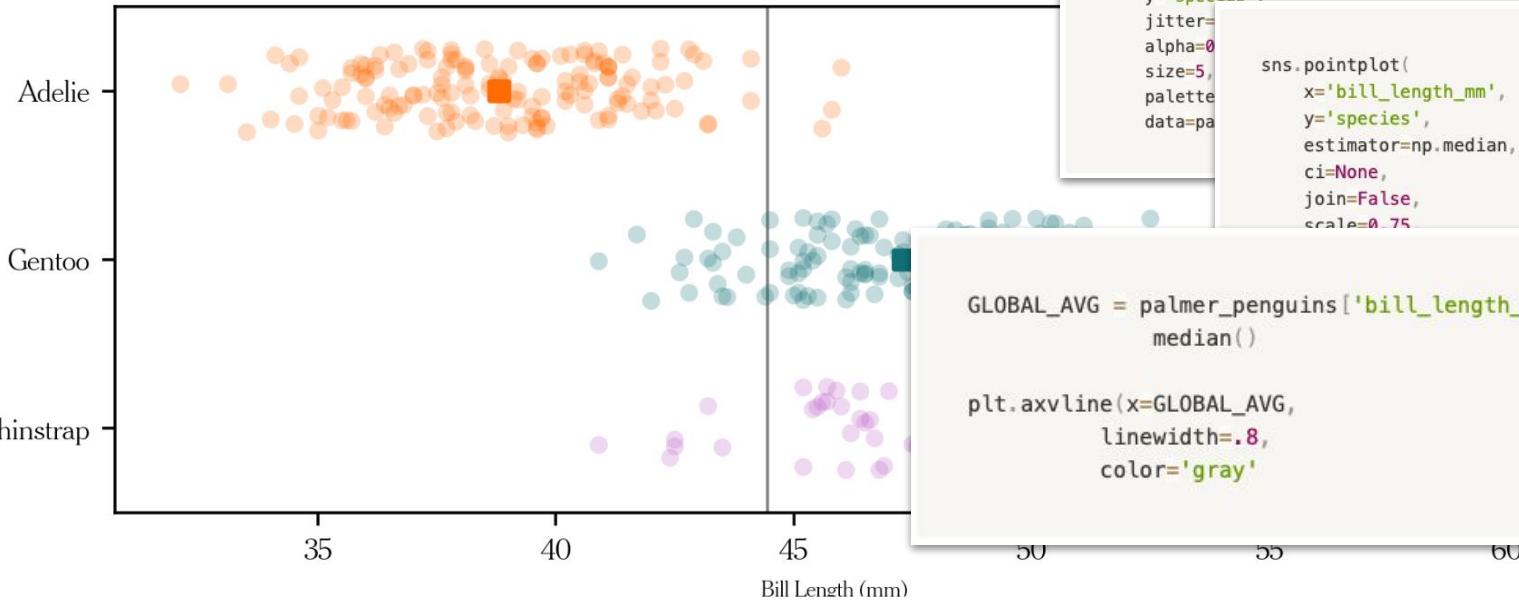
2.C Agregar otra capa de debajo agrega la información

```
[62]: 1 palmer_penguins[['species', 'bill_length_mm']].\  
2 | groupby('species').\  
3 | aggregate(np.median)
```



(2)

2.C Agregar otra capa de visualización: línea horizontal con mediana global



```
(1)
custom_palette = ['#ff6c04', '#107075', '#bf5cc9']

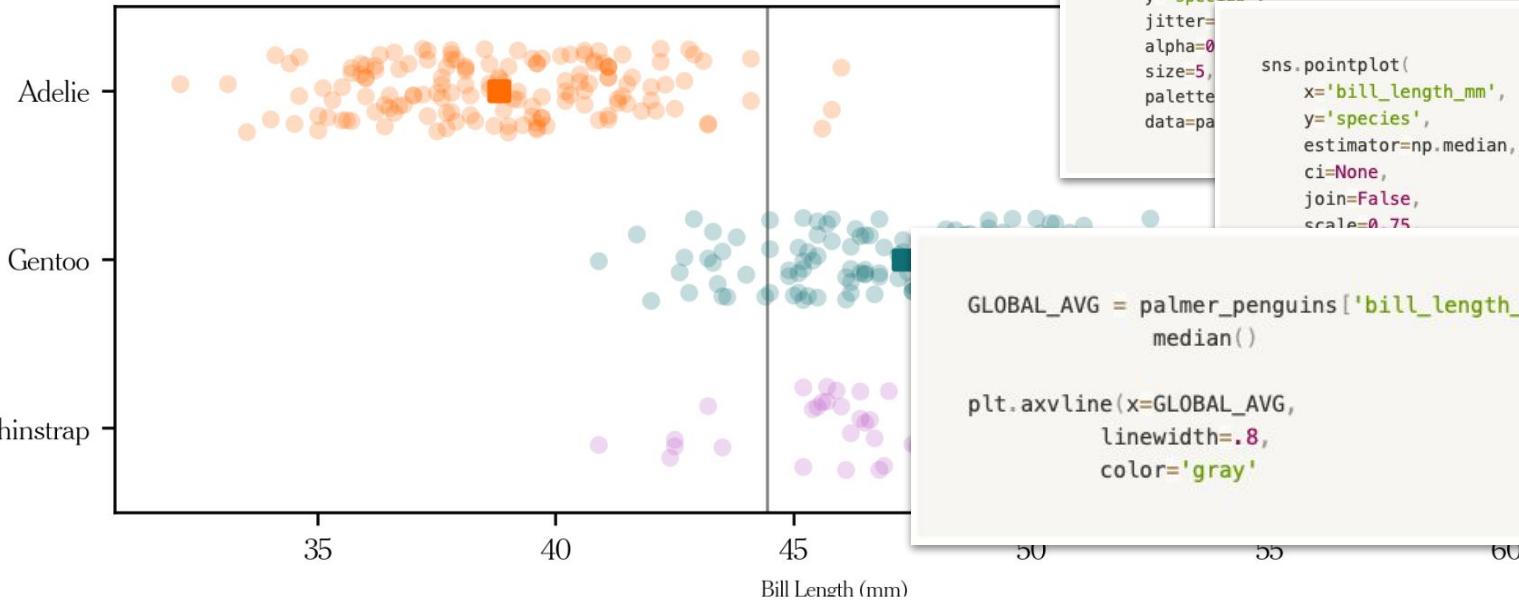
sns.stripplot(
    x='bill_length_mm',
    y='species',
    jitter=True,
    alpha=0.5,
    size=5,
    palette=custom_palette,
    data=penguins)
```

```
(2)
sns.pointplot(
    x='bill_length_mm',
    y='species',
    estimator=np.median,
    ci=None,
    join=False,
    scale=0.75)
```

```
(3)
GLOBAL_AVG = palmer_penguins['bill_length_mm'].median()

plt.axvline(x=GLOBAL_AVG,
            linewidth=.8,
            color='gray')
```

2.C Agregar otra capa de visualización: línea vertical representando la mediana global



```
(1)
custom_palette = ['#ff6c04', '#107075', '#bf5cc9']

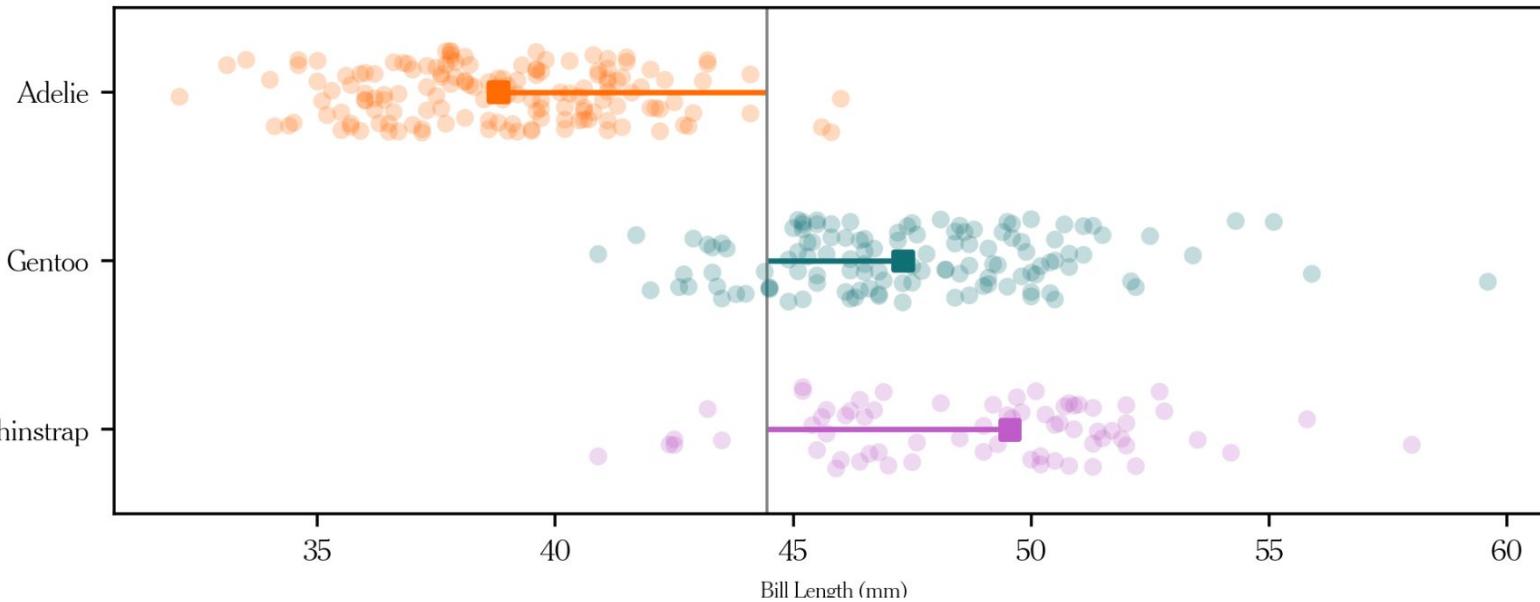
sns.stripplot(
    x='bill_length_mm',
    y='species',
    jitter=True,
    alpha=0.5,
    size=5,
    palette=custom_palette,
    data=penguins)
```

```
(2)
sns.pointplot(
    x='bill_length_mm',
    y='species',
    estimator=np.median,
    ci=None,
    join=False,
    scale=0.75)
```

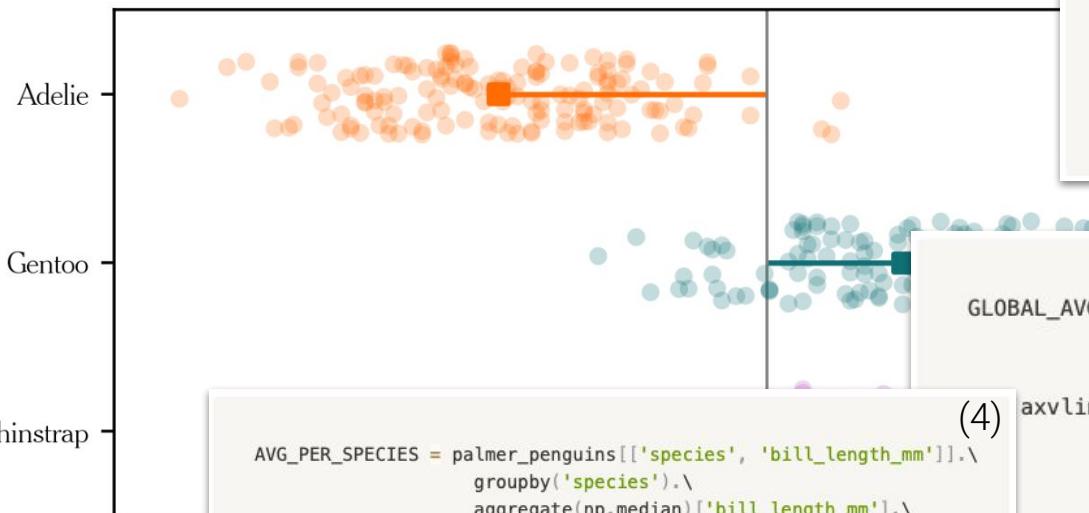
```
(3)
GLOBAL_AVG = palmer_penguins['bill_length_mm'].median()

plt.axvline(x=GLOBAL_AVG,
            linewidth=.8,
            color='gray')
```

2.C Agregar otra capa de visualización: distancia de la mediana global a la grupal



2.C Agregar otra capa de visualización: distancia de la mediana global a la grupal



(1)

```
custom_palette = ['#ff6c04', '#107075', '#bf5cc9']
```

```
sns.stripplot(  
    x='bill_length_mm',  
    y='species',  
    jitter=True,  
    alpha=0.5,  
    size=5,  
    palette=custom_palette,  
    data=penguins)
```

(1)

```
sns.pointplot(  
    x='bill_length_mm',  
    y='species',  
    estimator=np.median,  
    ci=None,  
    join=False,  
    scale=0.75)
```

(2)

(3)

```
GLOBAL_AVG = palmer_penguins['bill_length_mm'].median()
```

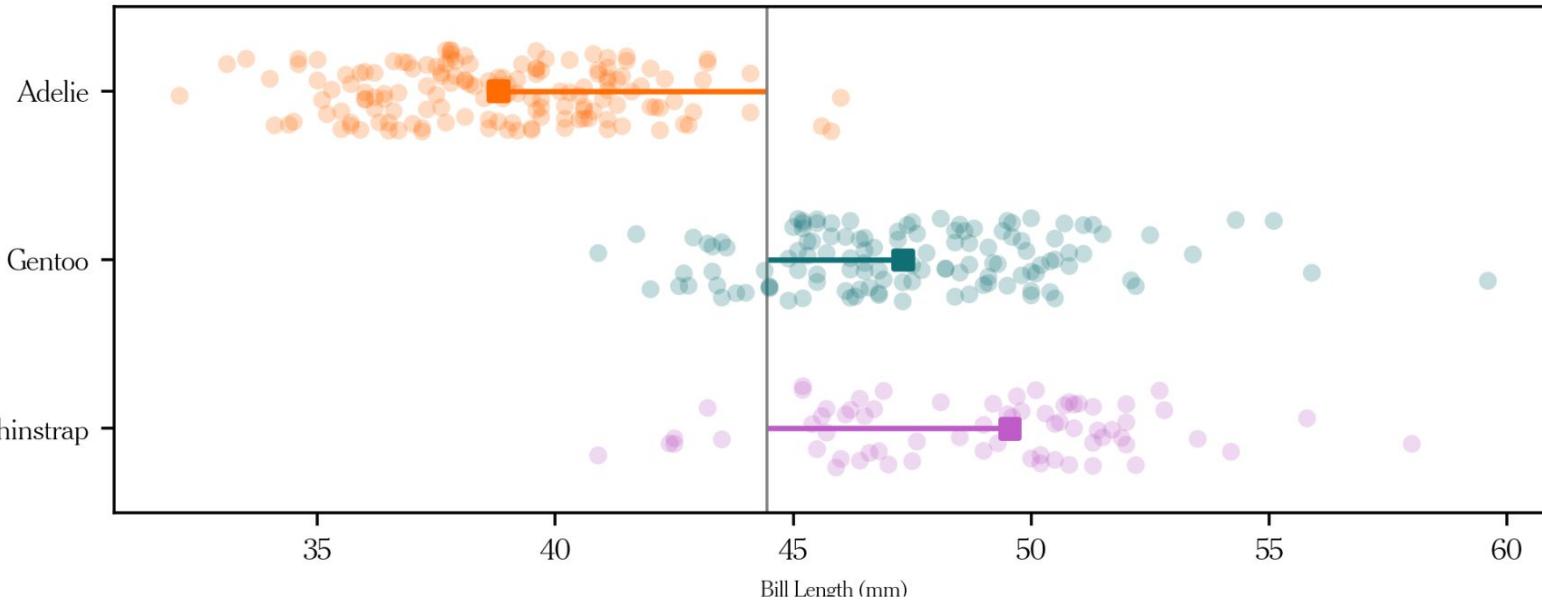
```
axvline(x=GLOBAL_AVG,  
        linewidth=.8,  
        color='gray')
```

(4)

```
AVG_PER_SPECIES = palmer_penguins[['species', 'bill_length_mm']].  
                    groupby('species').  
                    aggregate(np.median)['bill_length_mm'].  
                    sort_values()
```

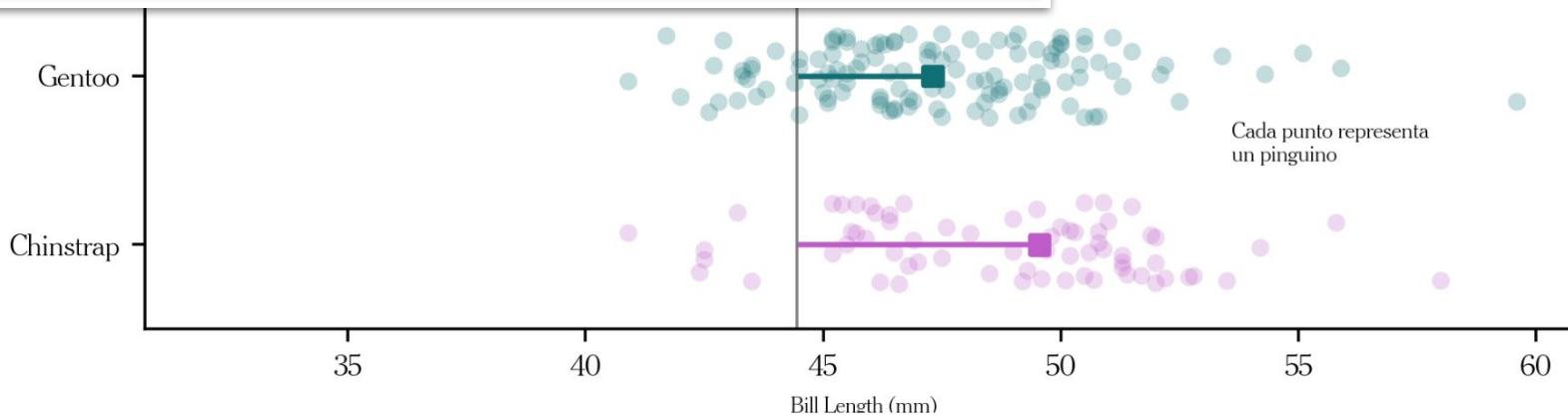
```
ax.hlines(y = palmer_penguins['species'].unique(),  
          xmin=AVG_PER_SPECIES,  
          xmax=GLOBAL_AVG,  
          color=custom_palette)
```

2.C Perturbaciones del stripplot son aleatorias: para lograr reproducibilidad se puede usar np.random.seed

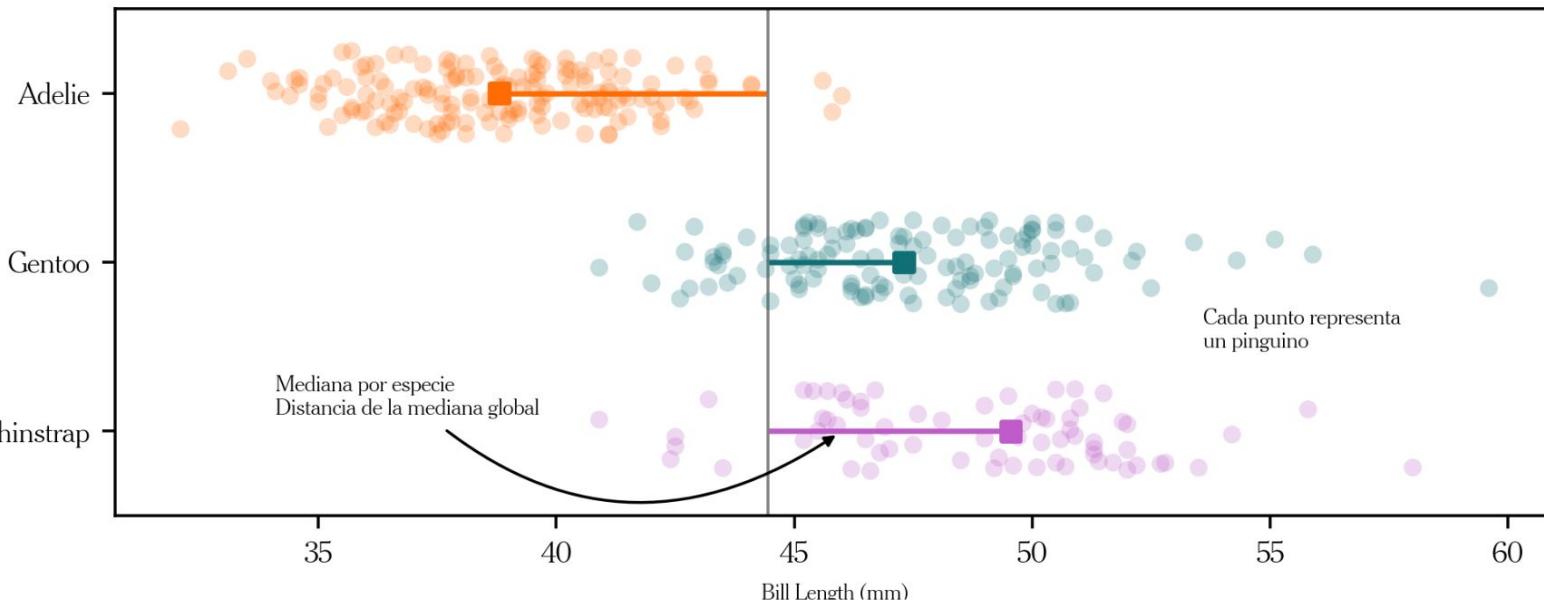


2.C Agregar anotaciones y flechas: utilizar ax.annotate

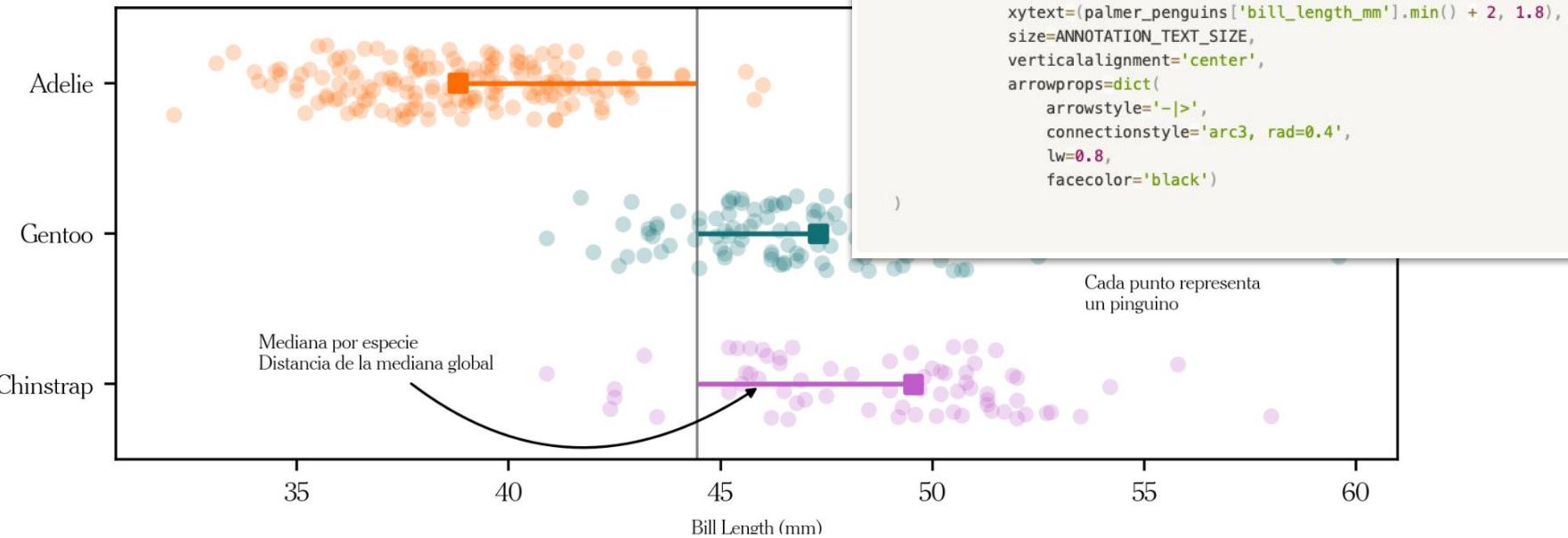
```
ax.annotate("Cada punto representa\nun pinguino",
            xy=(palmer_penguins['bill_length_mm'].max() - 6, 1.4),
            xytext=(palmer_penguins['bill_length_mm'].max() - 6, 1.4),
            size=ANNOTATION_TEXT_SIZE,
            verticalalignment='center')
```



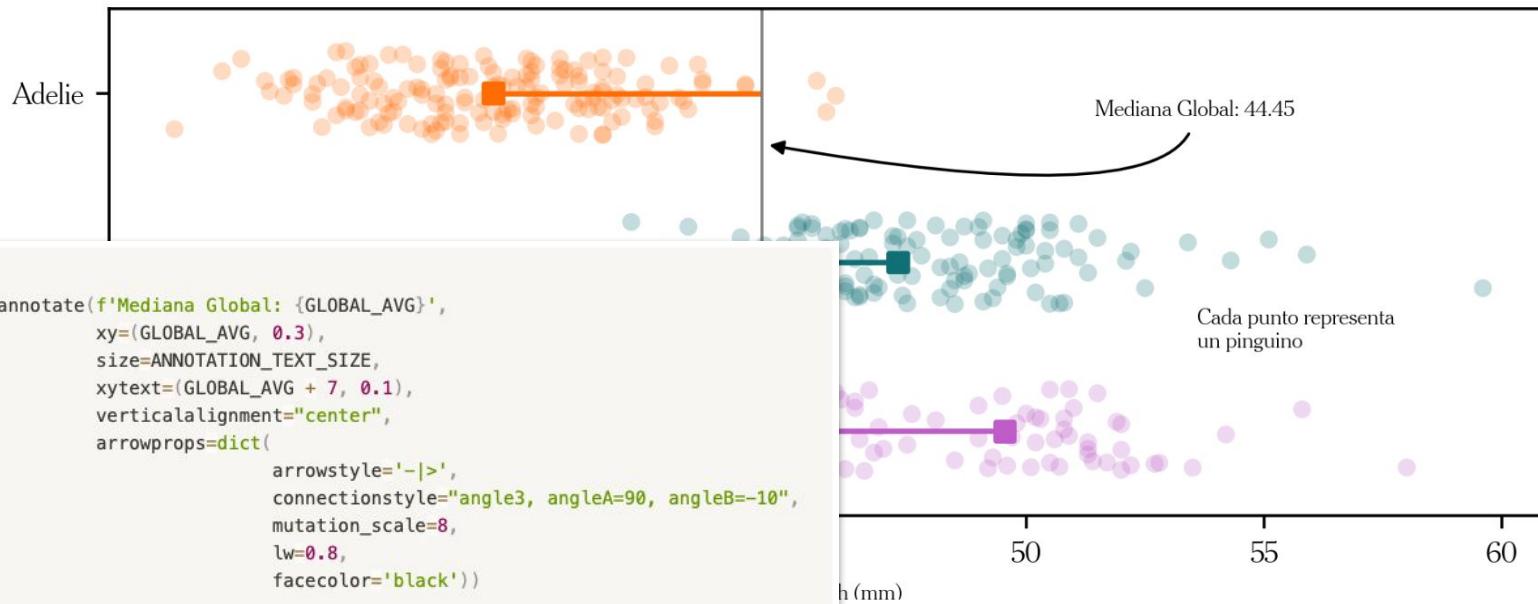
2.C Agregar anotaciones y flechas: dar contexto y explicar qué significan los elementos visuales



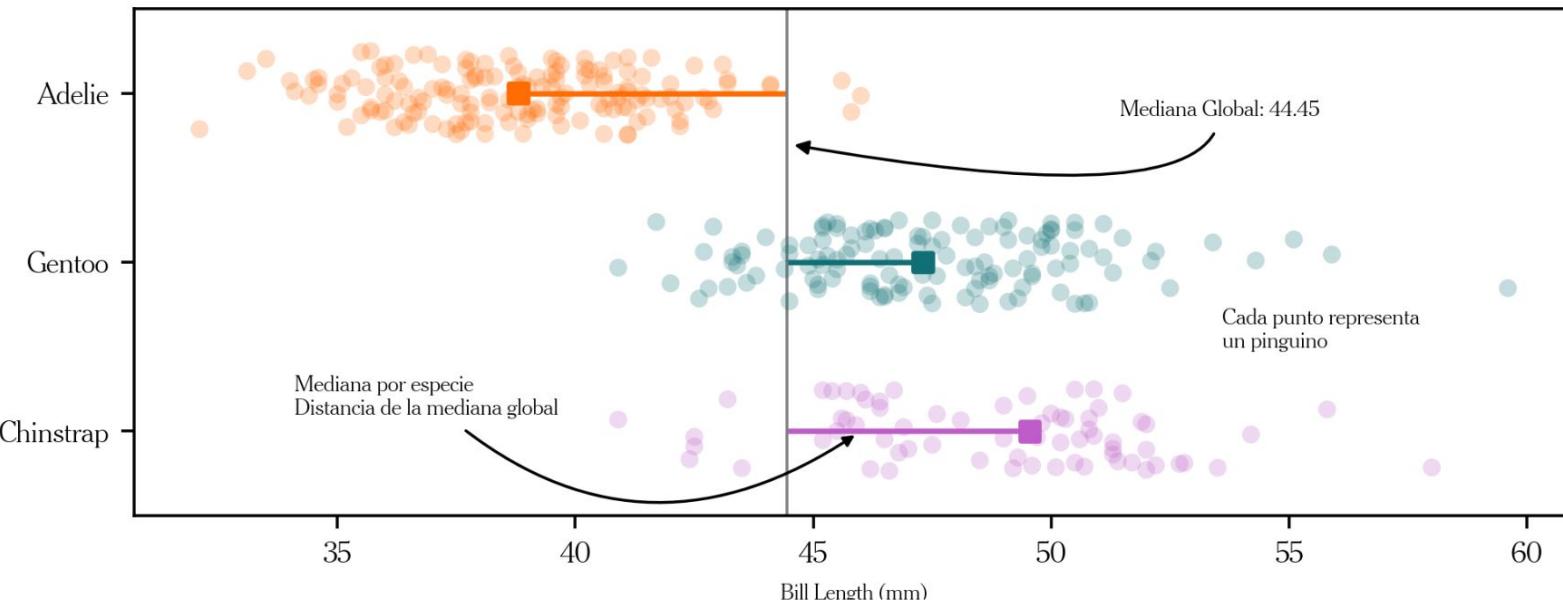
2.C Agregar anotaciones y flechas: dar contexto y explicar qué significan los elementos visuales



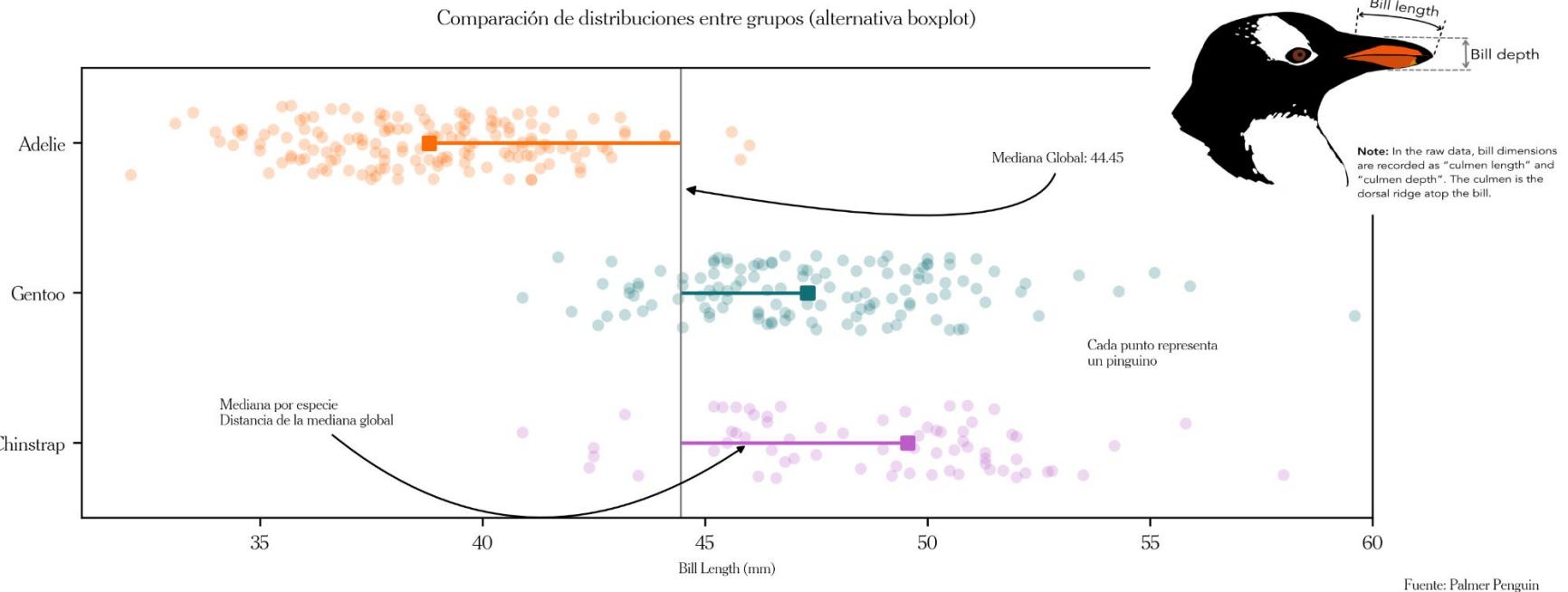
2.C Agregar anotaciones y flechas: dar contexto y explicar qué significan los elementos visuales



2.C Alternativa para comparar grupos: visualización por capas y anotaciones

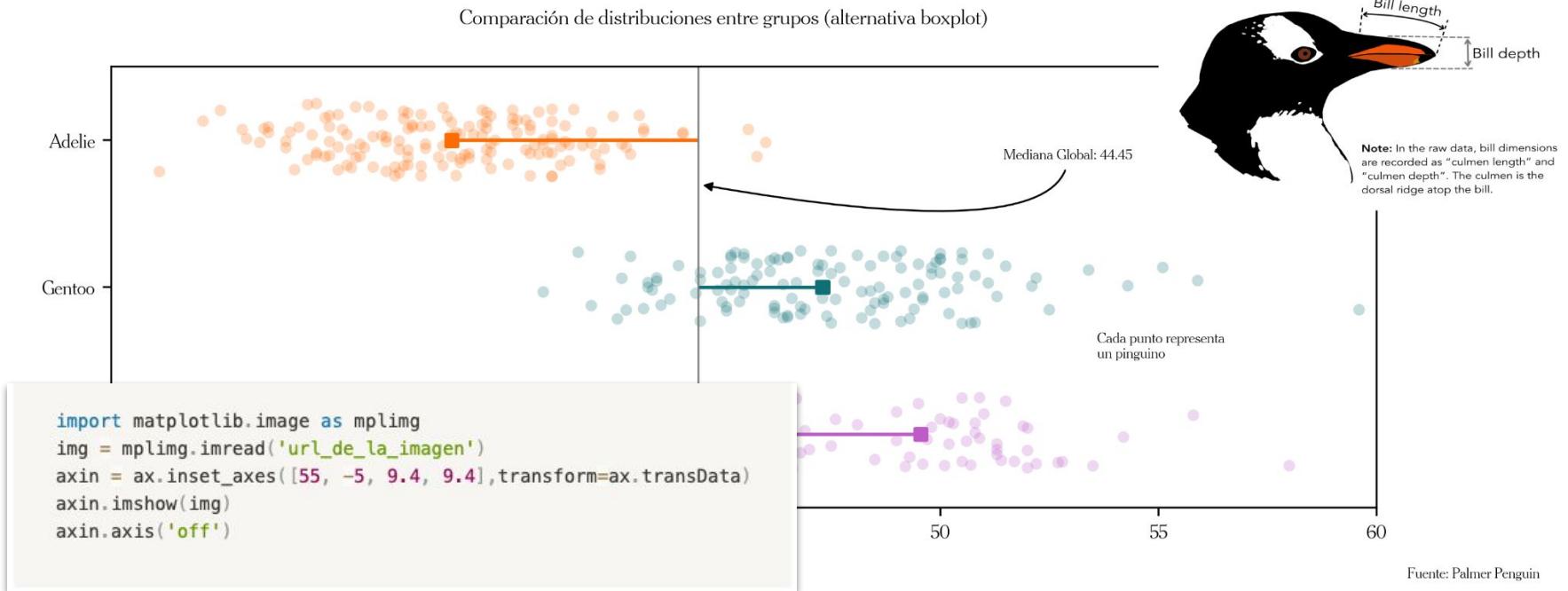


2.C Añade una imagen: agregar un axes dentro de otro con ax.inset_axes



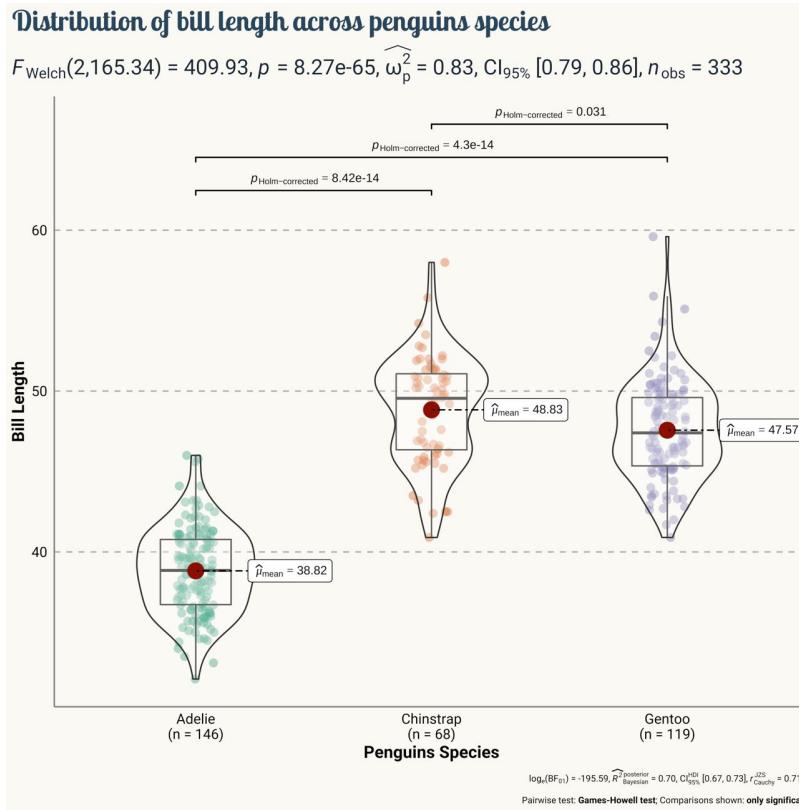
Recomendado: [The Evolution of a ggplot \(Ep.1\) \(Cédric Scherer\)](#)

2.C Añade una imagen: agregar un axes dentro de otro con ax.inset_axes



Recomendado: [The Evolution of a ggplot \(Ep.1\) \(Cédric Scherer\)](#)

2.D Otro ejemplo de visualización por capas y anotaciones



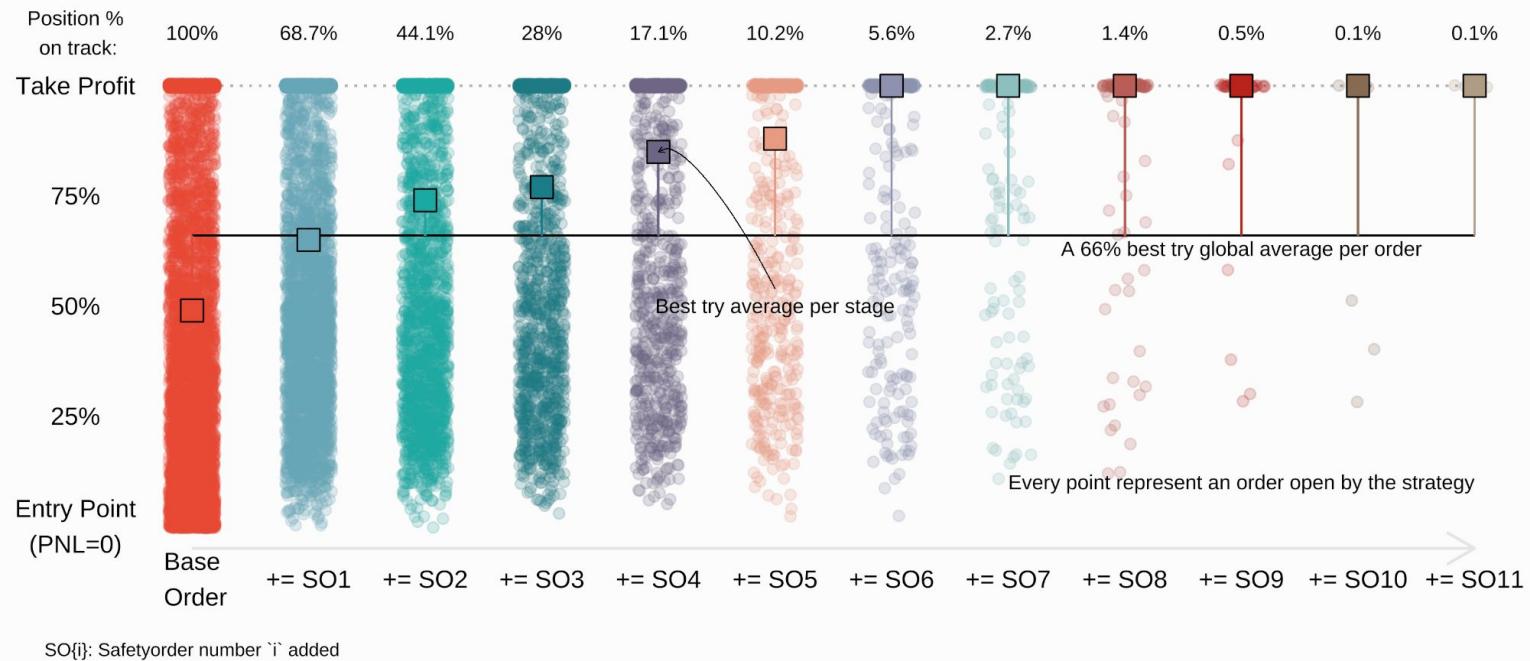
1. **Capas:** strip-chart + boxplot + violinplot + pointplot
2. **Considerar la audiencia:** a quién va dirigida la visualización?
 - a. Violinplot es menos intuitivo que un strip-chart
 - b. Si la audiencia es público general, los tecnicismos de los test-estadísticos pueden reducirse a "consideramos test-estadísticos y observamos diferencias significativas entre los grupos"

Recomendado: [Palmer Penguins exploration with violinplots in Matplotlib](#)

Safetyorder usage and completion profile

Best try value per order stage, value indexed between the entry and the take profit price

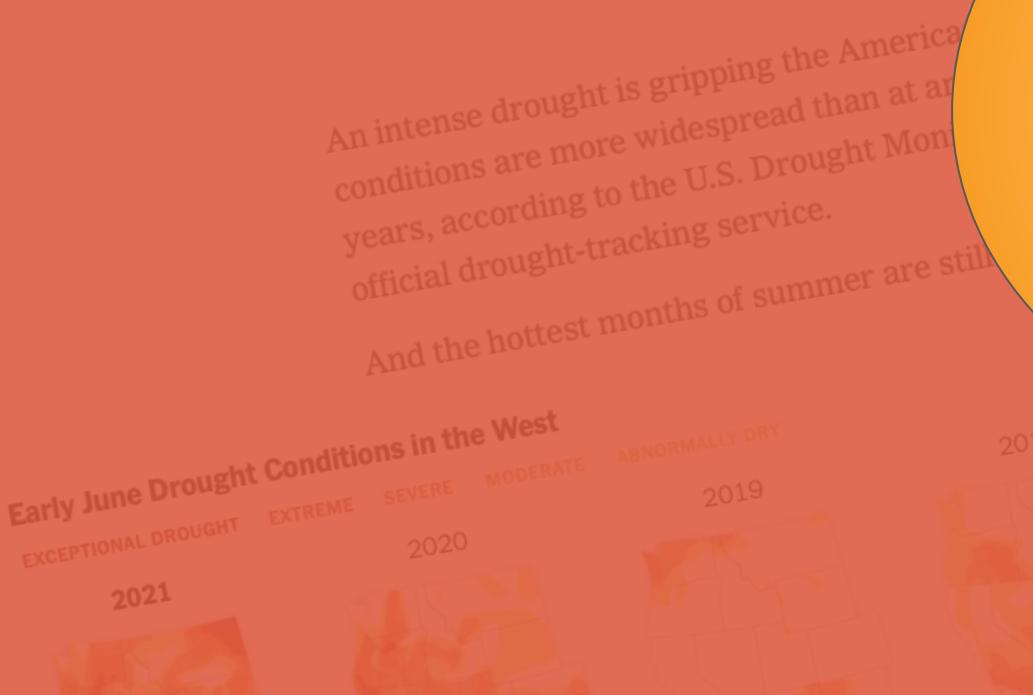
Every best try that reaches the take profit is a position closed that contribute to reducing the % position on track. Otherwise, the bot opens a new safety order



The New York Times

How Severe Is the Western Drought? See For Yourself.

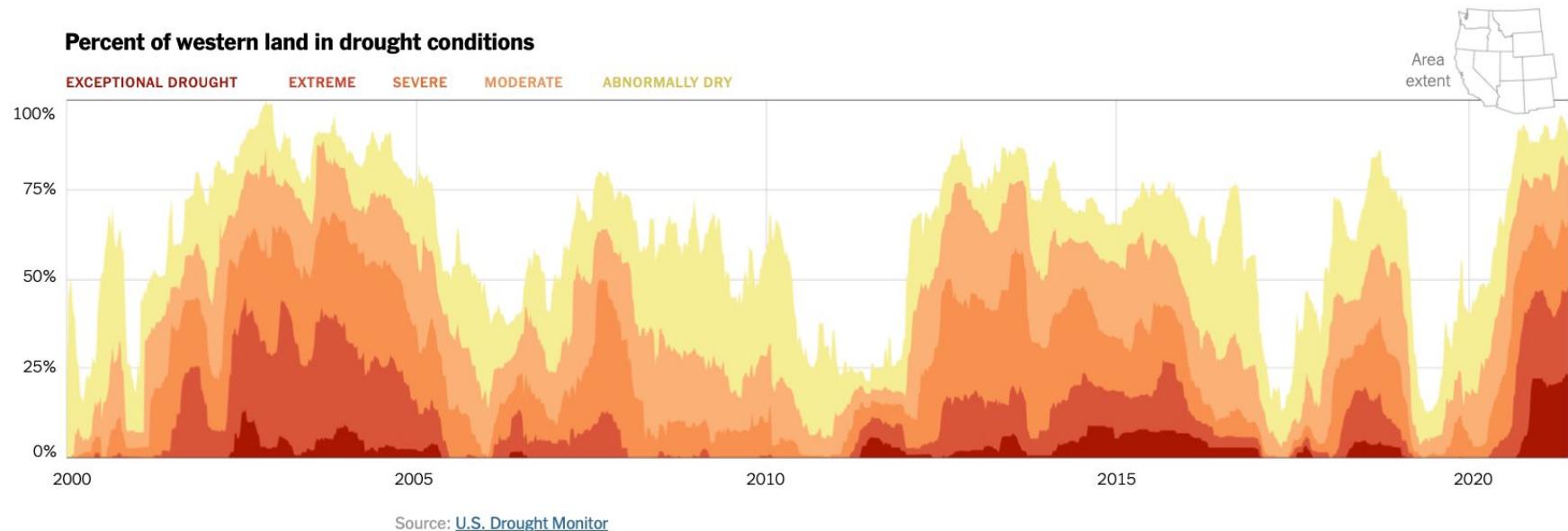
By Nadja Popovich June 11, 2021



#3

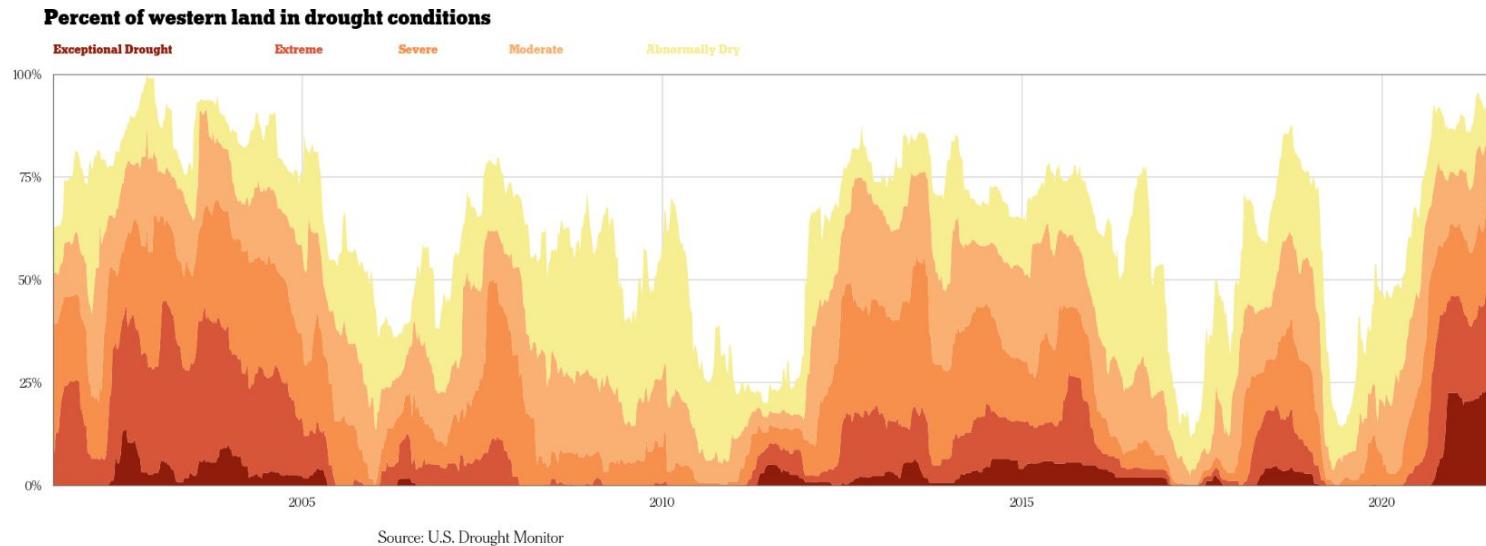
Caso U.S Drought

How Severe Is the Western Drought? See For Yourself.



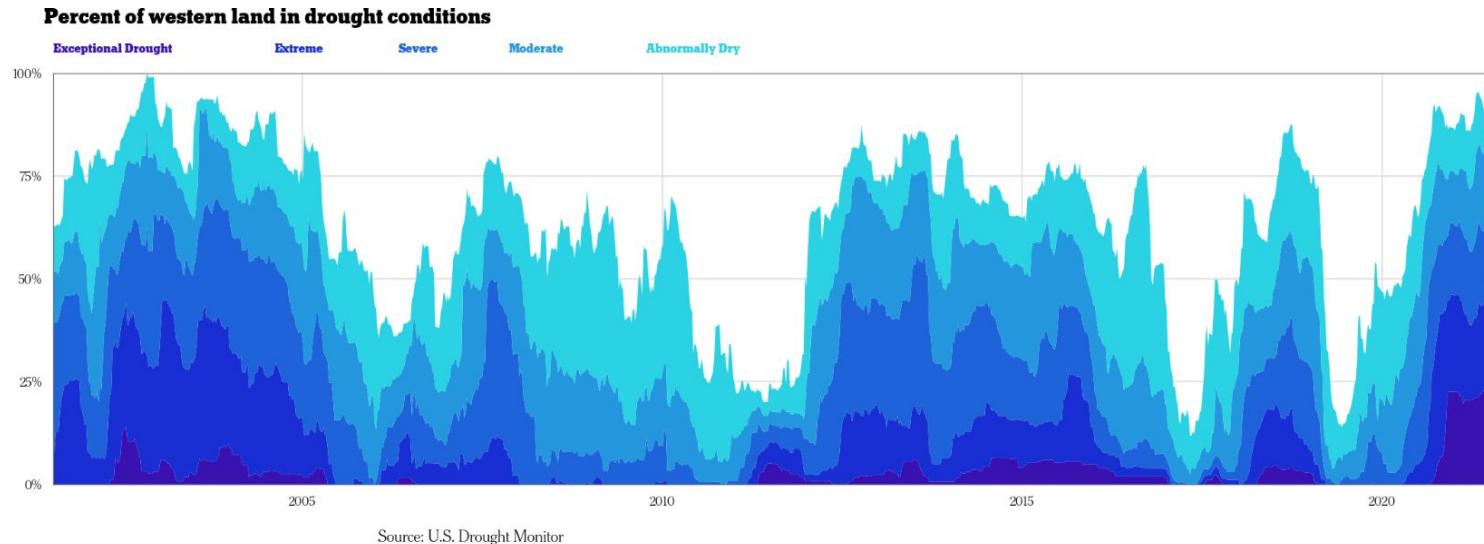
[The New York Times: How Severe is the Western Drought? See For Yourself](#)

3.A Replicar visualizaciones: guía para mejorar y adoptar buenas prácticas (importante lo que estás copiando)



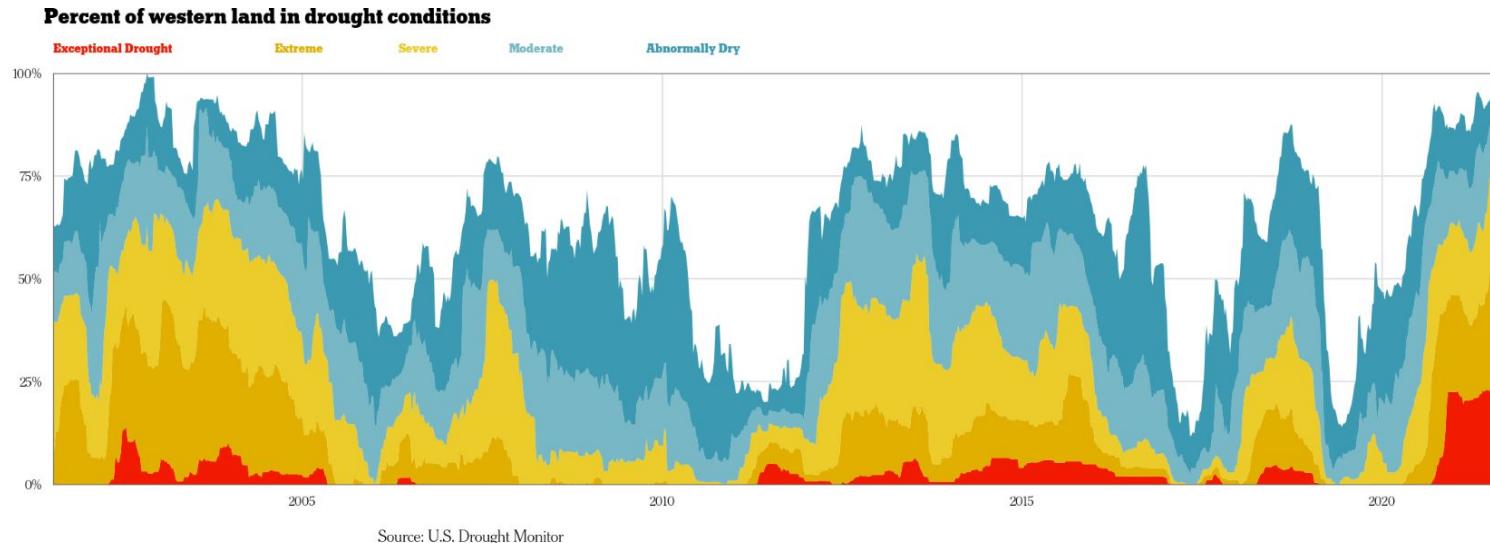
[The New York Times: How Severe is the Western Drought? See For Yourself](#)

3.B La importancia de los colores: ¿tienen relación con lo que está comunicando?



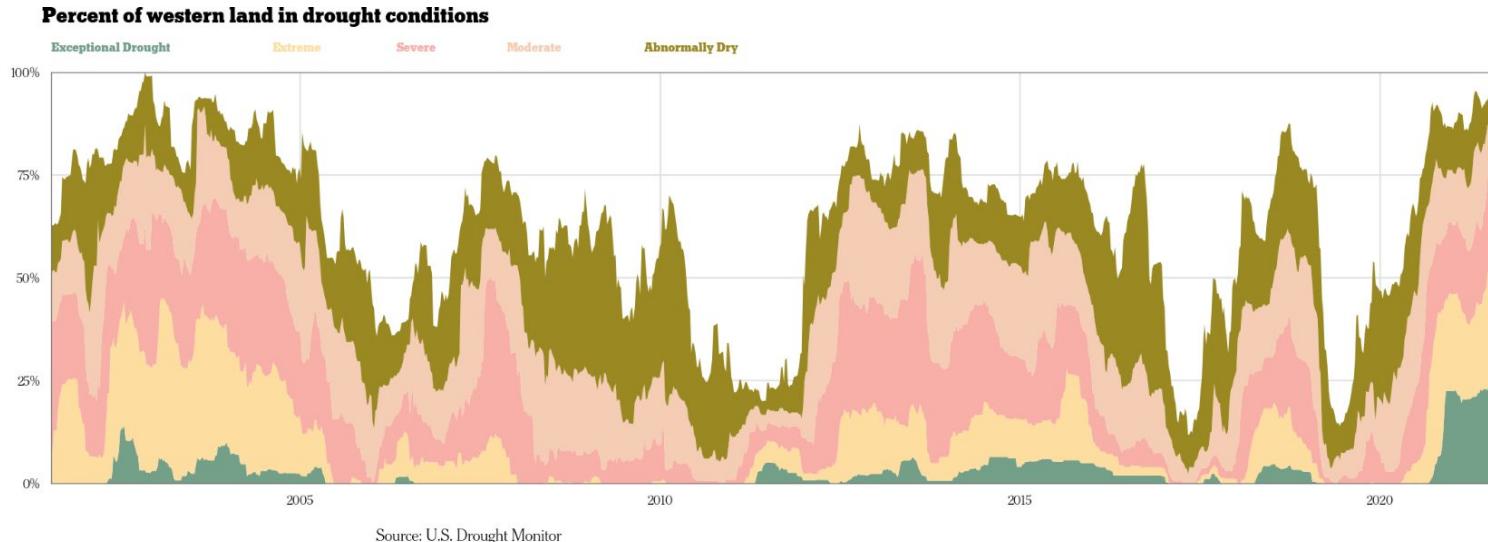
[The New York Times: How Severe is the Western Drought? See For Yourself](#)

3.C La importancia de los colores: rangos de colores de un extremo a otro



[The New York Times: How Severe is the Western Drought? See For Yourself](#)

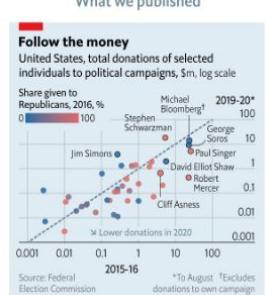
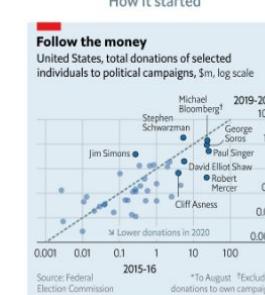
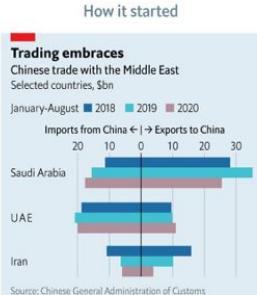
3.D La importancia de los colores: rangos de colores para categorías ascendentes/descendentes



[The New York Times: How Severe is the Western Drought? See For Yourself](#)

3.E La importancia de los colores: recomendaciones de lectura

- **Colors and Emotions in Data Visualization (Cédric Scherer)**: percepción del color y diferencias culturales en su significado
- **Off the Charts: How we use colour in our chart design (The Economist newsletter de periodismo de datos)**: usos del color en visualización de datos como tiempo, rango, para resaltar, y como data
- **Colores e Ilusiones**: monografía de la unidad de colores del curso de Visualización de Información por Eduardo Graells-Garrido





WOMEN IN DATA





Kevin Blake
@kevinsblake



Research tip: don't scroll through color palette databases hoping to find the perfect one for your fig.

Just copy the sprite of your favorite Pokémon into PowerPoint, select the eyedropper tool, and...



9:51 AM · Aug 24, 2021



♡ 8.5K ⚖ 56 ↑ Share this Tweet

Tweet your reply



Kevin Blake @kevinsblake · Aug 24, 2021



Research tip: don't scroll through color palette databases hoping to find the perfect one for your fig.

Just copy the sprite of your favorite Pokémon into PowerPoint, select the eydropper tool, and...



JUAN CRUZ LANDONI
@jclandoni

Update: there's a page for it! pokepalettes.com

11:43 AM · Aug 24, 2021



831



12



Share this Tweet

[Tweet your reply](#)

